



## CONTENTS

I.	INTRODUCTION .....	4
I.	A PORTABLE SPEECH PROCESSOR FOR INTERLEAVED-PULSES PROCESSING STRATEGIES	
A.	Interleaved-pulses processors as a general class .....	5
B.	General overview of hardware	
1.	Design criteria .....	13
2.	Specialized hardware .....	14
3.	Functional block diagram .....	18
C.	Front end hardware processing	
1.	Microphone input conditioning .....	22
2.	Spectral energy processing .....	22
3.	Squelch .....	24
4.	Automatic gain control .....	24
5.	Pitch extraction .....	24
6.	Battery monitor .....	25
D.	Post processing and software architecture	
1.	Software organization .....	26
2.	Spectral energy estimation	
a.	Basic computations .....	29
b.	Noise subtraction algorithm .....	31
c.	Detection of unvoiced speech energy .....	33
3.	Stimulus train generation .....	34
4.	Squelch and automatic gain control .....	41
5.	Pitch extraction .....	43
6.	Mode selection .....	50
E.	Stimulation output .....	51
F.	Power consumption .....	52
G.	Packaging .....	53
H.	Future improvements .....	62

III. PLANS FOR THE NEXT QUARTER ..... 63

References ..... 64

Appendix 1: Summary of Reporting Activity for this Quarter  
Appendix 2: "Comparisons of Processing Strategies for  
Multichannel Auditory Prostheses"  
Appendix 3: "Electrical Stimulation Model of the Auditory  
Nerve: Stochastic Response Characteristics"  
Appendix 4: "A finite-Element Model of Bipolar Field  
Patterns in the Electrically Stimulated  
Cochlea -- A two Dimensional Approximation"

## I. INTRODUCTION

The purpose of this project is to design and evaluate speech processors for auditory prostheses. Ideally, the processors will extract (or preserve) from speech those parameters that are essential for intelligibility and then appropriately encode them for electrical stimulation of the auditory nerve. Work in the present quarter included the following:

1. Evaluation of alternative processing strategies for single-channel auditory prostheses, in tests with a patient implanted with the extracochlear version of the 3M/Vienna device;
2. Continued psychophysical studies of loudness and pitch perception in an implant patient fitted with a percutaneous cable, as initially outlined in the Fifth Quarterly Progress Report for this project;
3. Development and testing of new computer programs to support and extend the above psychophysical and processor-evaluation studies;
4. Application and refinement of a portable speech processor in field trials with percutaneous cable patient MH;
5. Preparation of several manuscripts for publication (see Appendices 1 - 4); and
6. Presentation of project results in five talks (see Appendix 1), including one at the 1987 National Meeting of the Triological Society.

In this report we will describe in detail the design and application of the portable speech processor indicated in point 4 above. Descriptions of the studies indicated in points 1 and 2 above will be presented in future reports.

## II. A PORTABLE SPEECH PROCESSOR FOR INTERLEAVED-PULSES PROCESSING STRATEGIES

### A. Interleaved-pulses processors as a general class

Several major classes of speech processors for auditory prostheses have been evaluated in our laboratory over the past several years. Of these, one processor class, the interleaved-pulses processor, has emerged as a prime candidate for successful application in cochlear implant patients with either good or poor nerve survival. The basic interleaved-pulses strategy has been described previously in conjunction with preliminary studies with patient LP (QPR 7, NIH project N01-NS-2356). We also have applied this strategy to patient MH (QPR 2, NIH project N01-NS-5-2396) using our block-diagram compiler emulation software. More recently, a series of six patients, implanted with the 4-channel UCSF/Storz auditory prosthesis, have been evaluated with the interleaved-pulses strategy. Results for these patients with interleaved-pulses processors were compared with those for their existing compressed-analog processor. Results from this latter study will be presented in future quarterly reports.

The present report describes an initial version of a portable, real-time speech processor that implements the interleaved-pulses strategy. As described previously (QPR 3, NIH project N01-NS-5-2396), real-time speech processor hardware development is being pursued in three phases. Phase I is the development of a real-time, bench-level processor that will at minimum fully implement all of the features of processors simulated with the block-diagram compiler. This bench-level processor is intended as both a flexible research tool and as a prototype design for later portable processors. The phase I bench-level processor interfaces to the patient through our existing laboratory stimulation equipment. Phase II is a portable, hardware CHMOS implementation of the flexible processor system designed in Phase I. This portable processor will allow short-term evaluations of promising processor strategies by

patients in environments outside the laboratory. The physical package of the Phase II processor will be larger than final, permanent portable processors since minimization of package size and power consumption are secondary to functional flexibility at this design stage. The present Phase II portable processor interfaces to the patient via multiplying DAC's and voltage drivers which are memory-mapped on the microprocessor address bus. Phase III is the final development of the portable processor that the patient will use permanently. This design will be a highly optimized implementation of the best strategies identified using the previously described processors. Since many of the engineering decisions are dictated both by the information processing requirements of the processor design (processor cycle time, number of channels, number of bits) and by the constraints of achieving a final portable device (chip packaging, power requirements), all three development phases are being pursued simultaneously. Indeed, features of the present Phase II portable processor that have been added in response to the demands of noisy environments outside of the laboratory are now being retrofitted to the original Phase I bench-level processor.

This report describes the present status of the Phase II effort to develop a portable interleaved-pulses speech processor. Patient MH is currently using this device on a full-time basis away from the laboratory. It is worth noting at this point that patient MH utilizes this processor in an exceptionally noisy workplace environment. Speech plus noise to noise ratios are typically 2 dB (A-scale). Noise control measures to reduce the noise sources (water aerators with cooling compressors) reduce the noise approximately 6 dB. Various design changes have been implemented in the processor itself to minimize the remaining noise effects and will be discussed where appropriate.

Before proceeding, a brief review of the interleaved-pulses signal processing strategy is presented. Figure 1 shows a block diagram for an interleaved-pulses processor. In the interleaved-pulses processor an automatic gain control (AGC) continuously

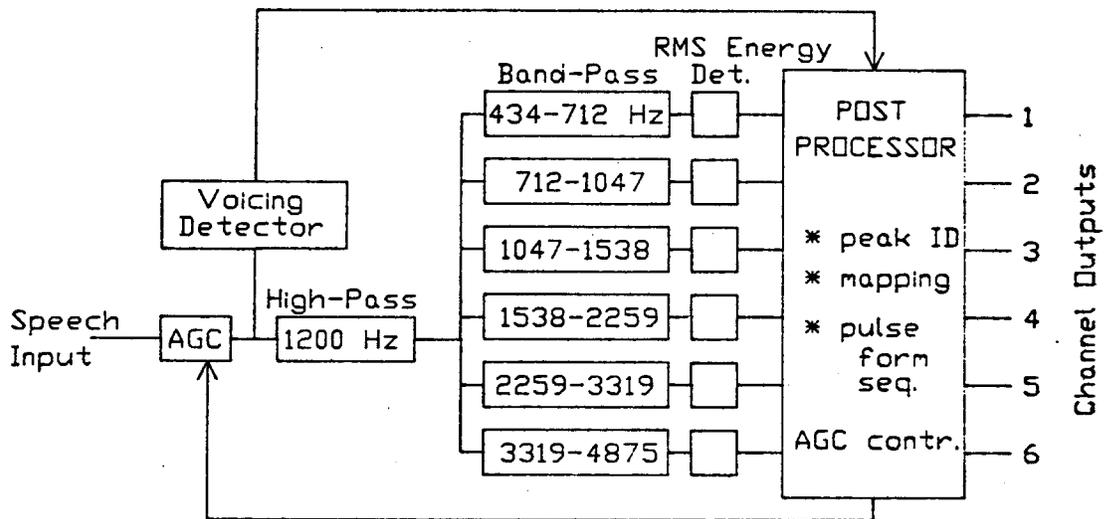


Figure 1. General functional block diagram of interleaved-pulses processor.

adjusts the level of speech input so that a steady average level is presented to subsequent stages of the processor. Typical attack and release times for the AGC are 16 and 320 msec, respectively, providing a "slow AGC" action. The level-adjusted signal is then high-pass filtered to reduce the amplitudes of speech components below 1200 Hz. The output of the high-pass filter is fed to a bank of band-pass filters whose center frequencies span the combined range of the first and second formants of speech, along a logarithmic scale. The root-mean-square (RMS) energy in each band is sensed by a full-wave rectifier and a low-pass filter connected in series to each band-pass filter output. Next, a "post processor" is programmed to scan the RMS outputs on a periodic basis. The output of a filter bank channel is coded for stimulation of its assigned electrode(s) only if the RMS energy is above a preset "noise threshold." The amplitudes of the pulses delivered to the selected channel(s) are determined from a logarithmic mapping law of the form:

for RMS level  $\geq$  to RMS<sub>thres</sub>

pulse amplitude =  $A \times \log(\text{RMS level}) + k$ ,

otherwise,

pulse amplitude = 0

where the parameters "A", "k" and "RMS thres" have been specified for each channel according to the threshold, most-comfortable loudness (MCL) level and noise level for that channel. Finally, the voicing detector senses the fundamental frequency of voiced speech sounds and whether a given speech input is voiced (periodic) or unvoiced (aperiodic). The output of the voicing detector can optionally be used by the post processor to control the timing of "round-robin" update cycles, as described below.

Variations of interleaved-pulses processors may be produced

through different choices of parameters for the post processor. These parameters include (1) the number of channels stimulated on each stimulus cycle; (2) the duration of stimulus pulses for each channel; (3) the interval between pulses on sequentially stimulated channels; (4) the order in which channels are to be stimulated; (5) the mapping law for each channel, as described above; (6) the waveforms of stimulus pulses; and (7) whether stimulus sequences are to cycle continuously or are to be timed according to information provided by the voicing detector. Parameters 1 through 4 define the basic sequence of stimulation across channels, which we term as one "round-robin" cycle. Round-robin cycles typically are repeated as rapidly as possible if voicing information is not to be explicitly coded. Alternatively, inputs from the voicing detector can be used to time the beginning of each round-robin cycle. If voicing information is to be explicitly coded, round-robin cycles are timed to start in synchrony with the fundamental frequency (F0) during voiced speech sounds and at either randomly-spaced or maximum-rate intervals during unvoiced speech sounds. Explicit coding of voicing information might be expected to improve a patient's perception of prosodic features associated with F0 contours and to help the patient make voice/unvoice distinctions for consonants (e.g., improve the ability to distinguish an "s" from a "z" or a "t" from a "d"). Also, an explicit representation of voicing information might be expected to improve the "naturalness" of speech percepts and the ability to make man/woman/child distinctions.

Typical waveforms for an interleaved-pulses processor are shown in Figure 2. In the figure the top trace is the input to the processor and the remaining traces are channel outputs. The input is the word "BOUGHT." The initial consonant occurs at about 180 msec and the vowel follows immediately thereafter. An expanded display of waveforms well into the vowel is shown in the lower-left panel. The "t burst" of the final consonant begins slightly before 640 msec, and an expanded display of waveforms

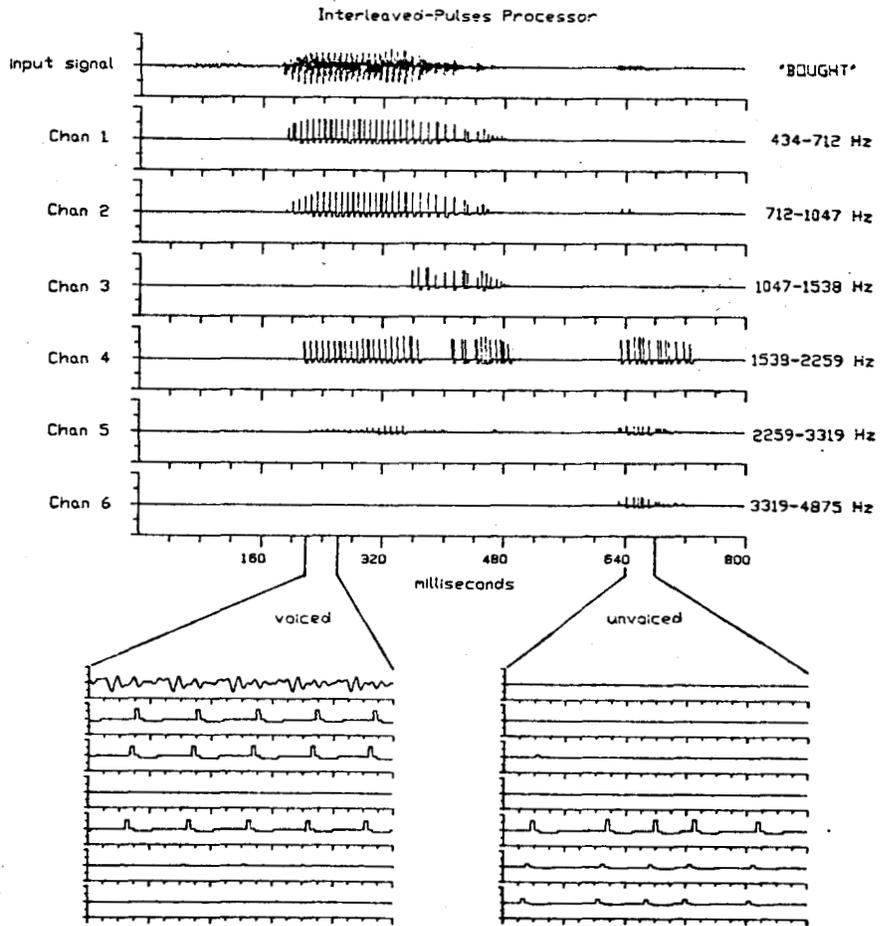


Figure 2. Typical waveforms for an interleaved-pulses processor showing output across several round-robin cycles.

beginning at 640 msec is shown in the lower-right panel of the figure. A striking feature of the interleaved-pulses processor is the sparseness of stimulation resulting from the use of non-simultaneous stimuli. In the particular variation of the processor presented in Figure 2, the greatest 4 of 6 channels are updated on every round-robin cycle and voicing information is explicitly coded. During voiced speech sounds the round-robin cycles are timed to begin in synchrony with the detected fundamental frequency, while during unvoiced speech segments the cycles are initiated at randomly-spaced intervals. The periodicity of cycle updates can be seen for a voiced speech sound in the lower-left panel of Figure 2 and the randomly-spaced cycle updates can be seen for an unvoiced speech sound in the lower-right panel. As mentioned before, the amplitudes of the pulses reflect the RMS energy levels in each channel's frequency band. Thus the timing of round-robin updates codes F0 for voiced speech sounds and also indicates whether a given speech sound is voiced or unvoiced. The speech spectrum above F0 is coded by the amplitudes of stimulus pulses and by the selection of channels. Many other variations of interleaved-pulses processors are available through manipulations of the parameters for the post processor.

The previous paragraphs have described the class of interleaved-pulses processors as it was originally conceived and implemented using the RTI block-diagram compiler software. Subsequent sections describe the implementation of this class of processor in a portable, real-time instrument. In moving from the high precision, low noise, and nonreal-time compiler emulation to an 8-bit precision, real-time processor, operating on noisy signals, several engineering compromises were required. These compromises will be discussed where appropriate.

## B. General overview of hardware

### 1. Design criteria

The Phase II portable speech processor is an intermediate step between the bench-level, real-time processor and the final optimized portable processor. In one sense it is the prototype for the final processor but it is also a research tool for testing advanced speech processor designs in the context of the daily activities of the patients. The basic design goals, stated roughly in order of priority, include:

- (a) full implementation of interleaved-pulses processing strategies, including those that code explicit voicing information;
- (b) broad flexibility that accommodates a wide range of processor designs, both parametrically and architecturally;
- (c) functional approximation to what would be expected for a final portable design;
- (d) electrical specifications including, a minimum of 8 bits of stimulus resolution, 100 usec temporal resolution, current driver outputs with + and - 10 volts voltage compliance;
- (e) small package size to be consistent with definition of a portable unit;
- (f) low power consumption so that battery size, weight and charging schedule do not interfere with daily use by patient;
- (g) ease and simplicity of operation for the patient; and

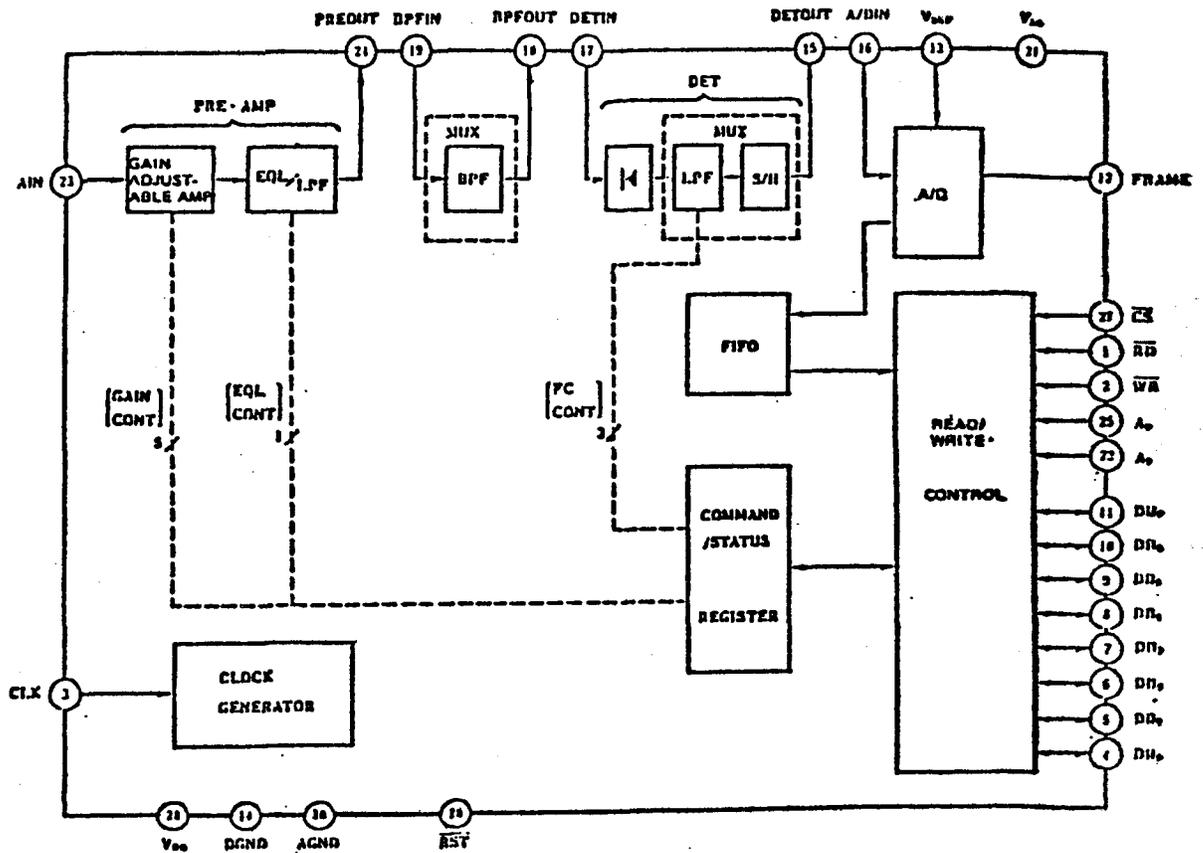
- (h) rapid design and production to make processors available to patients and to broaden the base of experience for the design of more effective speech processors as quickly as possible.

Of course, safety concerns are of extreme importance. At all levels of hardware and software design and construction, safety and reliability have been emphasized.

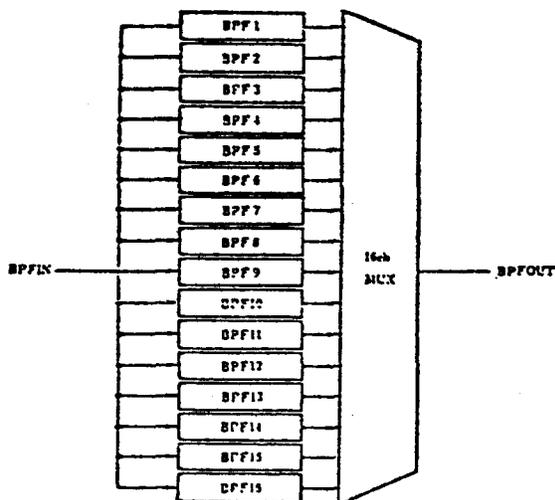
## 2. Specialized hardware

Only in recent years has the state-of-the-art of electronics developed to a point where portable, battery-operated implementation of complex signal processing strategies, such as the interleaved-pulses processor, has been feasible. This is largely due to development of large scale integrated circuits which utilize fast, low-power CMOS technology. Two devices incorporating these advances were selected for this first implementation. One is a microprocessor-compatible speech analysis chip for speech recognition systems that features a band-pass filter bank with RMS outputs. The other is a high speed programmable microprocessor that implements the post processor logic. Each of these devices is described briefly below.

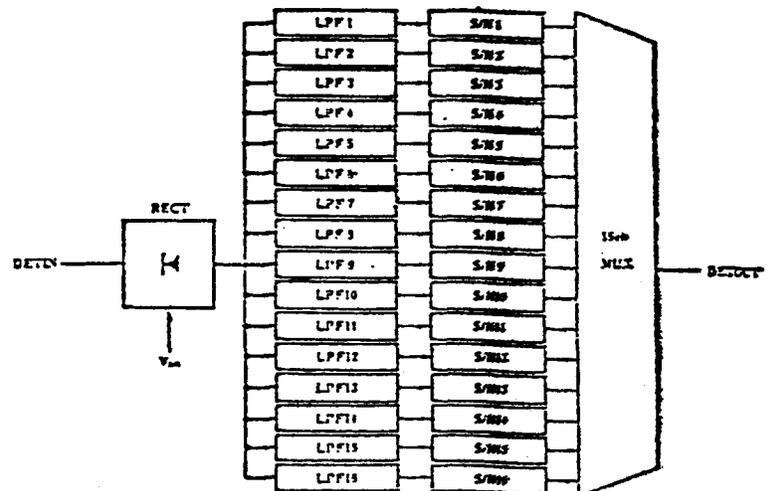
Filter bank analysis with RMS outputs is achieved with a single uPD7763D NEC speech analysis chip designed for speech recognition systems (Figure 3a). The uPD7763D incorporates a programmable pre-amp with optional equalizer, a 16-channel switched-capacitor band-pass filter bank (Figure 3b), a multiplexed rectifier with switched-capacitor low-pass filters and sample and hold (S/H) outputs (Figure 3c), and an 8-bit analog to digital (A/D) converter in a single LSI/CMOS 28-pin package. The uPD7763D has a general purpose microprocessor interface which provides access to a first-in, first-out (FIFO) buffer containing digitized RMS outputs of band energies. Pre-amp gain (-13.5 dB to +33.0 dB), equalizer ON/OFF, analyzed frame



(a) Block diagram of uPD7763D NEC speech analysis chip.



(b) Switched-capacitor filter bank with multiplexed output.

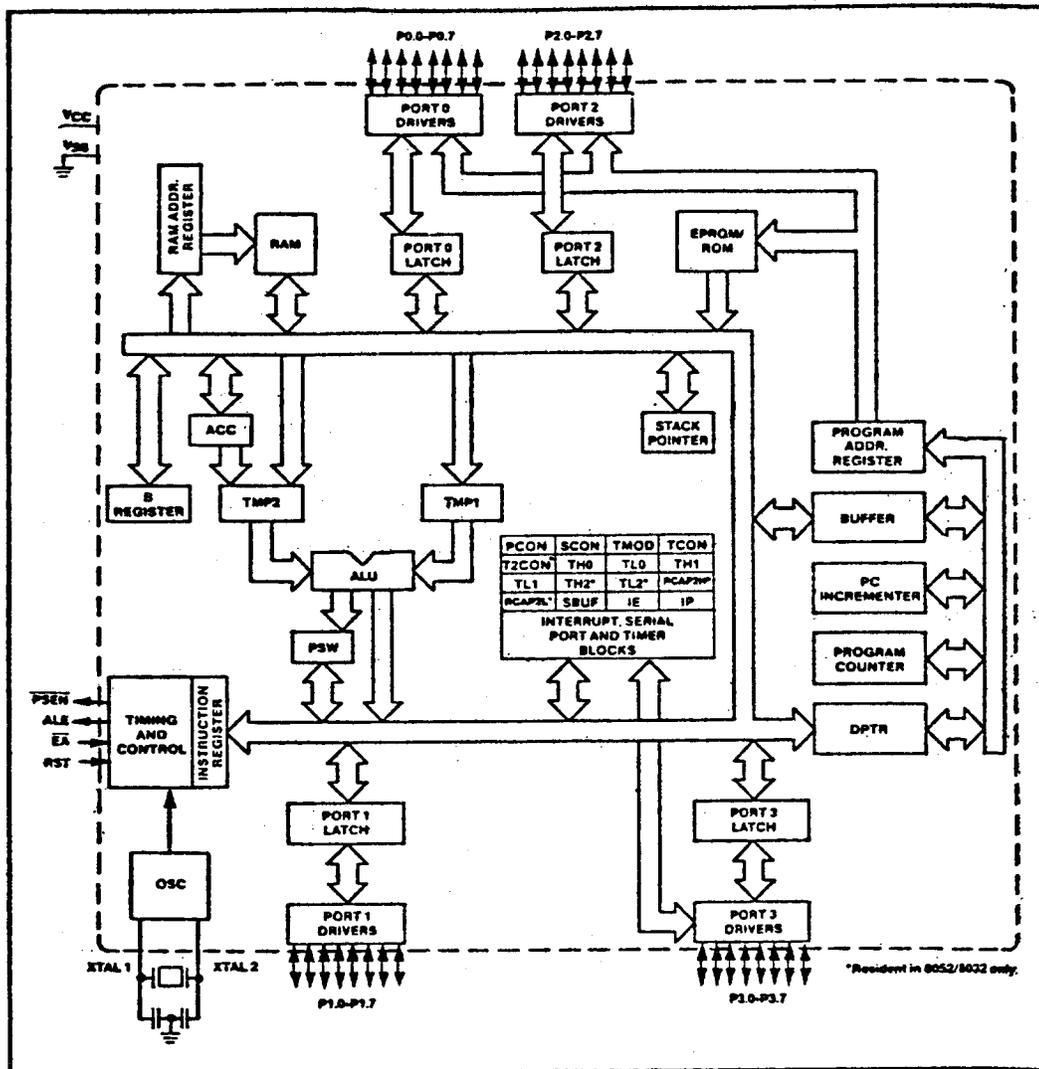


(c) Multiplexed RMS extractor featuring rectifier, low-pass filters and multiplexed S/H buffered outputs.

Figure 3. Functional components of NEC uPD7763D speech analysis circuit.

period (1-32 msec) and low-pass filter cut-off frequency (12.5 Hz - 400 Hz) are controlled via the microprocessor interface. Analysis proceeds on a frame-by-frame basis with RMS levels for each band-pass during the previous frame being available in a FIFO buffer. A frame period signal is available as an external interrupt signal to the post-processor, which may then read the memory-mapped FIFO buffer. The uPD7763D typically consumes 175 mWatts of power (350 mWatts maximum). This level of power consumption is higher than what would be desired for a final production unit. However, the uPD7763D circuit is still attractive for use in the initial portable instrument described here. It offers in a compact package a remarkably powerful set of speech analysis resources that would otherwise take considerable effort and expense to develop. Its power consumption is appropriate for a versatile portable speech processing instrument. Development of similar integrated circuits with lower power consumption may be premature until more is understood about the hardware requirements necessary for better speech processing strategies.

The post-processor is an Intel 8031 8-bit microprocessor (Figure 4). This device features two 16-bit programmable timers with interrupt capability along with two external interrupt lines. The 8031 has 128 bytes of onboard RAM and can address 64k bytes of external program memory and 64k bytes of external data memory. Thirty-two programmable I/O lines and an asynchronous serial port are available. A hardware multiplication and Boolean processing capability are also provided. This processor is available in CHMOS technology as the 80C31. The 80C31 additionally offers two software selectable power down/idle modes for minimization of power consumption. In the powerdown mode the processor clock is turned off, whereas in the idle mode the interrupt, asynchronous serial input, and onboard timer operations continue to function. At 12 MHz clock operation (1 usec instruction period, typical), the 80C31 consumes 80 mW of power at full operation and 22.5 mW of power when idled.



- Power Control Modes
- 128 x 8-Bit RAM
- 32 Programmable I/O Lines
- Two 16-Bit Timer/Counters
- 64K Program Memory Space
- High Performance CHMOS Process
- Boolean Processor
- 5 Interrupt Sources
- Programmable Serial Port
- 64K Data Memory Space

Figure 4. Block diagram of INTEL 80C31 8-bit microcontroller.

### 3. Functional block diagram

Figure 5 shows a functional block diagram of the present portable, real-time speech processor. Central to the design is the 80C31 microprocessor that functions as the postprocessor for the system. On a common data bus the controller can access 32k bytes of EPROM and 8k bytes of RAM memory, can both control the state of and read information from the NEC speech analysis chip, and can control eight D-to-A converters for output to the electrodes. Only six output channels are used in the present design for patient MH. In addition, the microprocessor handles signals to and from dedicated hardware circuits for the operation of the automatic gain control and squelch functions and samples running speech to identify glottal pulses during voiced intervals. Finally, the microprocessor monitors the battery status, senses the mode switch position and controls front panel light emitting diode (LED) indicators to show system status. Input signals from the microphone are initially filtered restricting them to speech spectrum frequencies. Outputs from the processor are transformer-coupled, voltage-controlled drivers instead of the more desirable current-controlled driver stages. The microprocessor operates at a 8.0 MHz clock rate.

The functional structure of both the general interleaved-pulses processor (Figure 1) and of this portable hardware implementation (Figure 5) may be divided into three rough stages of (i) front-end hardware processing followed by (ii) software-driven post processing culminating in (iii) output stimulation. The implementation of each of these stages in the portable processor is discussed in detail in the following sections. Generally the portable implementation closely follows the design of the interleaved-pulses processor as described in section II.A. There are however a few modifications that have been incorporated into the portable processor design to enhance performance and obtain the same functional design as described in section II.A. These modifications are discussed briefly here.

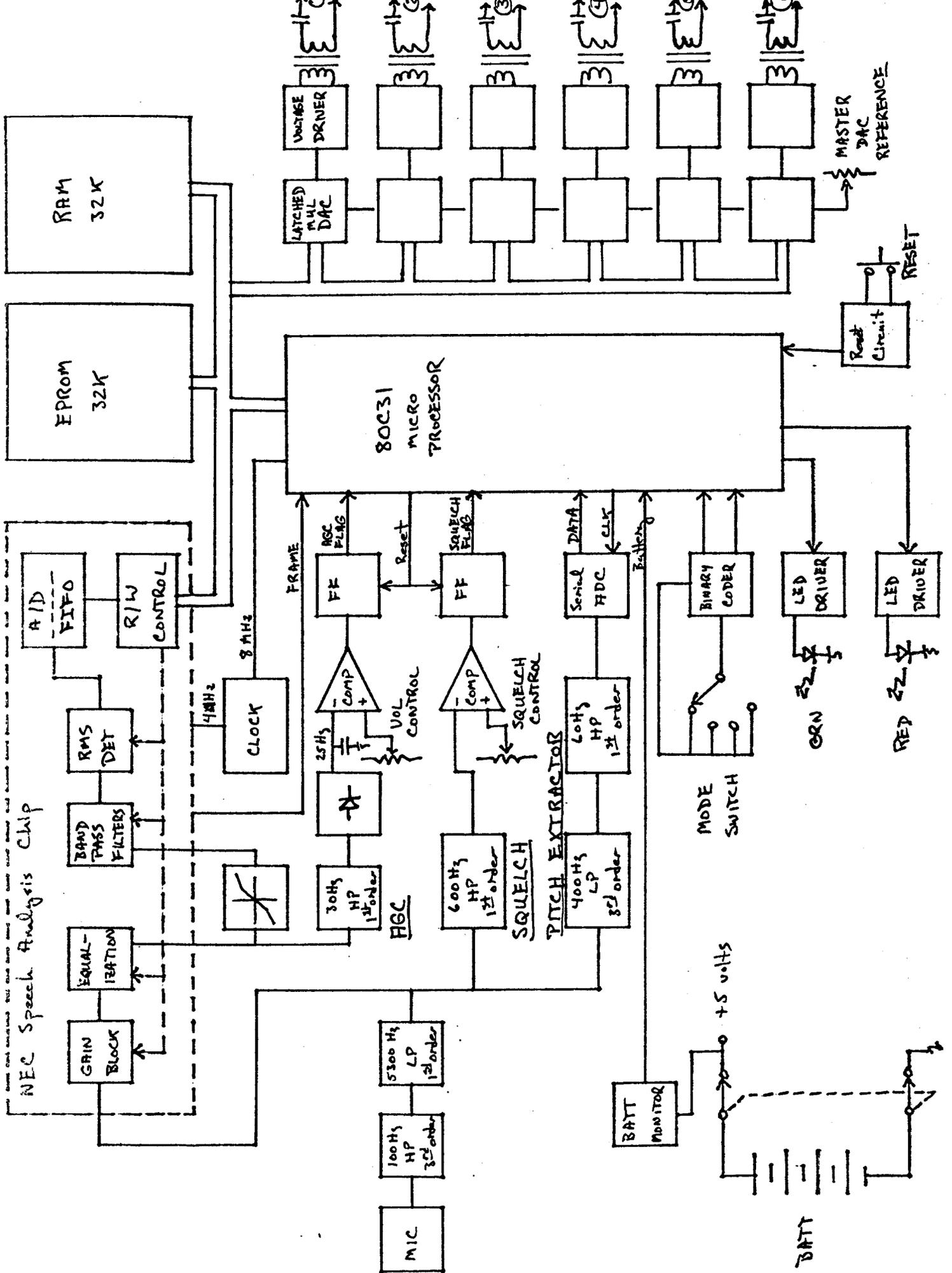


Figure 5. Functional block diagram of portable interleaved-pulses processor.

The first modification is that the front-end 1200 Hz. high-pass filter that follows the AGC has been implemented in software following the band-pass filter bank. This method adds flexibility in that almost any front-end filter characteristic can be easily specified in software. In addition, separate front-end hardware stages are eliminated thus saving space and reducing power consumption.

The second modification pertains to how stimuli are delivered during nonvoiced portions of speech. In the high accuracy (floating point), low noise block-diagram compiler implementation of the interleaved-pulses processor round-robin cycles are presented continually at either randomly-spaced or maximum-rate intervals during nonvoiced speech. In mapping RMS energy levels to pulse amplitudes,  $RMS_{thres}$  levels are selected so that during quiet passages output pulse amplitudes are zero. This is perfectly adequate for low noise speech, producing characteristically sparse output stimulation patterns like that of Figure 2. In the presence of noise however, continuous presentation of round-robin cycles during nonvoiced speech produces a continual low level of stimulation. When tested with such a stimulus, patient MH reports a smearing or fusion of the discrete speech sounds and consequently scores lower on confusion matrix tests. This phenomenon arises for two reasons. One reason is the higher noise levels associated with the portable processor due to both environmental sound sources and intrinsic electrical sources within the processor itself (i.e. front end hardware noise, internal NEC chip noise). The other reason is the decreased amplitude resolution associated with 8-bit integer arithmetic of the microprocessor as compared to floating point arithmetic of the block-diagram compiler. When scaled to 8 bits,  $RMS_{thres}$  levels are typically smaller than one least significant bit level, making low level RMS band energy mapping coarse at best. To correct this problem, a separate unvoiced speech energy detector was added in software to discretely signal when round-robin frame cycles should be presented during nonvoiced speech. Details of the unvoiced speech energy detector are described in

section II.D.2.c. The net effect is that the portable, real-time processor outputs round-robin cycles of stimulation only during voicing or when substantial unvoiced speech energy is present. Otherwise, the processor presents no output. This resolves the background noise problem and restores the round-robin cycle structure to its original sparse representation of the speech signal.

The third modification is the addition of a squelch circuit to the processor. This squelch circuit is adjusted by the patient so that noise and speech plus noise intervals can be discriminated by the processor. This adjustment is easily done with the patient using her own voice as a test signal. Once set the squelch is rarely adjusted, unless the patient moves into a different environment where the background noise level is substantially different. The squelch circuit makes several useful features possible. One is that it serves as a coarse indicator of when the processor should and should not deliver stimuli to the electrodes. Another is that it allows the microprocessor to exploit its idle down power conservation features during squelched periods to extend battery life. This particular feature has not yet been fully implemented in the portable processor. The final feature the squelch function allows is described in the following paragraph.

The fourth modification also pertains to a noise associated problem. As stated earlier, patient MH utilizes her portable processor in a noisy workplace environment. We have implemented a noise subtraction algorithm which utilizes the processor's squelch circuit to discriminate noise from signal plus noise conditions. The noise is relatively constant compared to the time scale of speech and can be reasonably described by RMS energy estimates during times when the squelch is active. Subtraction of these RMS noise levels from the running speech plus noise RMS estimates improves the performance of the portable speech processor. This feature is discussed in greater detail in section II.D.2.b.

## C. Front end hardware processing

### 1. Microphone input conditioning

The microphone is a miniature electret device with a frequency response of 50-15000 Hz. Section II.G on packaging describes placement of the microphone.

Signals from the microphone are first high pass filtered at 100 Hz with a 3rd order Butterworth filter, followed by a variable gain stage with a single pole 5300 Hz low-pass cut off. Signals at this point are typically 3.10 V peak-to-peak for loudly spoken voiced sounds.

### 2. Spectral energy processing

After input conditioning, signals are fed directly to the NEC speech analysis chip. The first processing stage of the NEC device is the programmable gain block. The gain of this stage is set under software control and may vary from -13.5 dB to +33.0 dB. This gain stage is the forward path gain control element for the AGC control loop. Further details of the AGC are discussed in sections II.C.4 and II.D.4 below.

Following the gain stage is an optional, software-selectable equalizer. Equalization provides essentially a +6 dB/octave gain slope across the speech spectrum. The equalizer option is utilized in this processor design.

The output from the equalizer is then compressed instantaneously through a piece-wise-linear gain stage. Input-output characteristics for this gain stage are:

$$E_{out} = 2 * E_{in} \quad \text{for } E_{in} < 0.4 \text{ volts}$$

$$E_{out} = E_{in} + 0.4 \quad \text{for } E_{in} > 0.4 \text{ volts.}$$

This provides an instantaneous compression of high amplitude speech signals and reduces the crest factor (peak to RMS ratio) of the speech signal from 7 to 5. This compression effectively expands the dynamic range of the band-pass filters by allowing higher RMS inputs but avoiding peak clipping. In addition, considerable improvement is made in the signal to noise ratio through the band-pass filters since RMS levels are higher relative to the noise intrinsic to the NEC speech analysis chip itself.

Signals next are routed to the NEC chip's switched-capacitor, band-pass filter set. Here the speech spectrum is divided into sixteen bands, corresponding roughly to adjacent critical bands for normal human hearing. The output from each filter is then rectified and low-pass filtered (50 Hz, single-pole) to produce RMS energy estimates. These RMS energies are then sampled and the values are stored in the FIFO array. This array may be accessed from the microprocessor data bus and appears as a single memory mapped location. Sequential reads from that location pass the band energy information to the microprocessor. Once the FIFO has been loaded, the NEC chip sets the FRAME line to indicate to the microprocessor that more recent spectral information is available. Frame periods are 8 msec long.

In terms of the overall system functional block diagram (Figure 1), it is important to note that the 1200 Hz high-pass filter which reduces F0 and F1 energies is implemented in software. In the frequency domain, this filter produces staggered scaling of the outputs of the band-pass filters which is readily executed in software during reading of the band-pass energies by the post processor. By achieving the desired filter characteristic in this way, a greater degree of flexibility is gained, implementing more complex front end filters without any change in front end hardware.

### 3. Squelch

Squelch operation is a simple matter of determining whether or not a signal exceeds a preset threshold level. Speech signals from the microphone preconditioning stages are high-pass filtered (600 Hz, single pole) to diminish F0 energy. The resultant signal's positive phase is then compared to a reference dc level which is set with the front panel squelch control. If the signal level exceeds the reference then a logical flag is set. This logical flag is mapped into the bit space of the microprocessor via port 1 and is readily examined by the software. The microprocessor can issue a signal that clears the squelch flag, (simultaneously clearing the AGC flag as well). The software logic of the squelch function is described in section II.D.4., on squelch and AGC processing.

### 4. Automatic gain control

The AGC control loop operates to maintain the average RMS energy of the forward speech signal path near an operating point set by the front panel volume control. The output from the gain/equalization stage of the NEC speech analysis chip is half-wave rectified and low-pass filtered (25 Hz, single-pole). This RMS estimate is then compared to a DC level set by the volume control. If the RMS estimate exceeds the reference level a logical flag is set. This flag is bit mapped into the microprocessor via port 1 and is easily interrogated. It may be cleared along with the squelch flag, by the microprocessor. The software logic of the AGC is described in section II.D.4.

### 5. Pitch extraction

Pitch extraction is implemented essentially as a peak-picking software algorithm. At this front end hardware stage, running speech is preprocessed by filtering and then sampled. The microphone preconditioned signal is low-pass filtered (400 Hz, 3rd order Chebyshev, 0.5 dB ripple) to reduce F1 resonance

ringing. The signal is then high-pass filtered (60 Hz, single-pole) and finally digitally sampled at a 2.0 kHz rate using a serially interfaced analog-to-digital converter (ADC), TLC548. The ADC communicates with the microprocessor using the 80C31's on-board serial interface operating in mode 0. Chip select for the ADC is controlled by a port 1 line from the microprocessor.

## 6. Battery monitor

The battery voltage monitor is the voltage monitoring circuit built into a Maxim MAX 630 DC-to-DC converter. Reference levels are adjusted so that if the battery terminal voltage drops below 4.5 volts a logical line is cleared to the microprocessor. The microprocessor monitors this line as part of its background processing and will stop speech processing and signal the patient when the battery drops below 4.5 volts. At that point the processor stops all stimulation of the electrodes and begins flashing the red front panel LED at half-second intervals indicating that a battery change is required.

## D. Post processing and software architecture

### 1. Software organization

Post-processor functions for the six-channel, interleaved-pulses processors include:

- read the RMS energy estimates for the sixteen frequency bands from the filter bank FIFO levels;
- appropriately scale RMS energy estimates as implementation of the front-end, 1200 Hz high-pass filter;
- condense these sixteen band energies down to the required six band energies for the six-channel processor;
- adjust the band energy estimates for the presence of background noise in the RMS energy levels;
- determine if unvoiced speech energy is present;
- sort the six band energies into rank order by energy;
- sort the four maximum energy ranked channels into specific base-to-apex order for round-robin frame stimulation order;
- construct round-robin frame output buffer information by defining temporal and amplitude stimulus features for each channel from a previously-defined lookup table;
- output the current round-robin frame information based on the presence of voiced or unvoiced energy and the squelch circuit status;
- perform pitch extraction doing peak-picking detection of voiced energy;
- service the automatic gain control by continuously adjusting the forward path gain of the NEC chip;
- service the squelch hardware to determine if a signal is present;
- monitor the mode switch position and modify the operating mode if a change has been made;
- and, monitor battery status and disable speech processing if battery voltage is too low.

Post-processor software organization has been structured to take advantage of the hardware interrupt and timing features of the 80C31 microprocessor. Two independent timers are used to drive interrupt service routines which operate at different priority levels. Figure 6 summarizes the software organization and subsequent sections describe each software activity in more detail.

At first (highest) priority is the Stimulus Output Interrupt Task driven by timer 0 at 100 usec intervals. This task controls the loading of the output channel DACs with appropriate stimulus level information at appropriate times as defined by the round-robin frame buffer. Software flags from the pitch extractor and the unvoiced speech energy detector are used to determine when to initiate a round-robin cycle. The DAC buffers are latched, permitting the service routine to encode only changes in the output data, thereby greatly reducing software overhead and freeing computing resources for lower priority tasks.

At second priority is the Pitch Extraction Interrupt Task driven by timer 1 at 500 usec intervals. This task performs peak picking of F0 frequency information to detect pitch pulses. If a pitch pulse is detected, the Stimulus Output Interrupt Task is informed by the setting of a software flag.

All remaining processing is done on a background, non-interrupt basis. After an initial setup phase to establish the interrupt structure and preset required variables, the background task operates in a fast loop, checking for (1) a frame signal from the NEC chip, (2) mode switch position changes and (3) battery status. If a NEC frame signal is detected, the NEC Service Routine is called which (1) reads current spectral information from the NEC chip, (2) corrects for background noise presence in RMS energy estimates, (3) determines if unvoiced speech energy is present, (4) computes the round-robin frame information and stores it in an external RAM buffer, (5) services the squelch circuit, (6) services the AGC loop, and (7) at the

SOFTWARE ORGANIZATION

Start up

- set stack pointer
- zero DACs
- do self check
- indicate start-up LED pattern
- setup timer 0
- setup timer 1
- setup NEC chip
- setup AGC
- preset miscellaneous flags

Background Task

- check for FRAME signal from NEC  
if so, call NEC Service Routine
- check MODE switch for status  
if change, invoke new  
operating mode
- check for BATTERY flag  
if present, branch to  
shutdown  
otherwise, loopback

Shutdown Routine

- stop all outputs
- cycle RED LED on and off  
at 0.5 sec intervals  
indicating battery  
discharge state

Stimulus Output Interrupt Task

Priority 1 interrupt driven by timer 0  
(100 usec service interval, 10 kHz)

- checks Round Robin Frame timing and  
drives channel DACs with information in  
Round Robin Frame Data Buffer in  
internal RAM

Pitch Extraction Interrupt Task

Priority 2 interrupt driven by timer 1  
(500 usec service interval, 2 kHz)

- performs peak picking of F0 frequency  
information to detect pitch pulses

NEC Service Routine

effectively Priority 3, called by  
Background Task after FRAME signal from NEC  
chip is detected (8 msec interval)

- computes current spectral information
- makes voice/unvoiced interval  
determination
- orders all 6 channels by energies
- orders max 4 energy channels by channel
- computes pulse heights, channel nos.,  
duration and real-time timing for  
next Round Robin Frame and loads data  
buffer in external RAM
- services squelch circuit
- services AGC loop
- waits for current Round Robin Frame  
to end and then transfers external  
RAM buffer to internal RAM buffer
- return

Figure 6. General software organization for post processor.

appropriate time loads the new round-robin frame information into an internal RAM buffer for fast output by the Stimulus Output Interrupt Task. If a mode switch position change is detected, appropriate mode changes are executed, such as parameter, flag, or jump command modifications, and the background task is resumed. If the battery is discharged, then all output processing is immediately halted and the red LED is continuously flashed to inform the patient.

## 2. Spectral energy estimation

### a. Basic computations

The basic computation for spectral energy estimation consists of adding together RMS energy estimates for various combinations of bands from the NEC speech analysis chip. As described previously, the NEC chip performs analysis across sixteen bands, spaced roughly as adjacent critical bands for normal hearing. However, the interleaved-pulses processors require energy estimates from a smaller number of logarithmically spaced bands. Figure 7 shows the -3 dB break frequencies for the sixteen NEC analysis bands that span the spectrum from 146 Hz to 5756 Hz. Band 1 (146 - 340 Hz) is not used in the spectral analysis since energy in this band is predominately fundamental voicing energy, F0. Therefore, only outputs from 340 Hz to 5756 Hz, bands 2 through 16, are used from the NEC chip. Figure 7 also shows the desired frequency breaks for speech processors using from two to eight logarithmically spaced bands between 340 and 5756 Hz. Shown in parentheses beneath the frequency breaks for each processor are the NEC bands combined to approximate each processor band. While the NEC band combinations do not produce exactly logarithmically spaced bands, the approximations are fully adequate.

NEC band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
-3 dB freq Hz	146	340	536	732	927	1146	1366	1560	1780	2049	2366	2683	3073	3512	4073	4878	5756
8 band processor	340	484	690	982	1399	1992	2838	4041	5756								
		(2)	(3)	(4)	(5-6)	(7-9)	(10-12)	(13-14)	(15-16)								
7 band processor	340	509	763	1143	1712	2565	3842	5756									
		(2)	(3)	(4-5)	(6-8)	(9-11)	(12-14)	(15-16)									
6 band processor	340	545	873	1399	2242	3592	5756										
		(2)	(3-4)	(5-6)	(7-10)	(11-13)	(14-16)										
5 band processor	340	599	1054	1856	3269	5756											
		(2)	(3-4)	(5-8)	(9-12)	(13-16)											
4 band processor	340	690	1399	2838	5756												
		(2-3)	(4-6)	(7-12)	(13-16)												
3 band processor	340	873	2242	5756													
		(2-4)	(5-10)	(11-16)													
2 band processor	340	1399	5756														
		(2-6)	(7-16)														
NEC band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	

Figure 7. NEC Bandpass Break Frequencies showing band combinations for eight to two band interleaved-pulses processors.

In combining the energy estimates across several NEC analysis bands, attention must be given to the possibility of roll-over of the eight bit registers during addition. Scaling of individual energy estimates prior to addition is presently being used to avoid roll-over and maximize speed. Double precision addition with post division is more accurate, but is computationally more involved and therefore slower. Future versions will probably use the latter approach.

#### b. Noise subtraction algorithm

A noise subtraction feature has been added to the processor design in direct response to patient MH's experiences with her portable processor. As stated earlier (section II.A), patient MH usually wears her processor in a noisy workplace environment. While environmental noise reduction measures have helped to improve the processor's performance, further noise immunity remains a highly desirable feature for the processor.

Since the background noise in this case is due to continuously operating machinery, its spectrum is essentially constant on the time scale of speech processing. It is therefore possible to subtract the pure noise spectral energy estimates from the speech-plus-noise spectral energy estimates to derive spectral energy estimates for speech alone. Pure noise signals are discriminated from speech-plus-noise signals by the squelch feature of the processor. We call this process our noise subtraction algorithm and have implemented it as a mode switch option for MH.

Implementation of noise subtraction in software occurs immediately after the logarithmically-spaced energy band estimates have been obtained. If the noise subtraction mode is selected and the squelch mode is active, the current RMS energy estimates are considered to be pure noise and the current energy values are stored in buffer memory space. Processing continues,

but no output is generated since the squelch is active. If the squelch mode is inactive, then the RMS energy estimates are considered to be speech plus noise. At that point, the previously measured noise energies, stored in memory when the squelch was last active, are subtracted from the current RMS band energies. The resultant energies are then passed through subsequent processing stages to generate output stimuli.

This noise subtraction approach, in addition to removing slowly-varying environmental noise contamination, will also remove processor generated steady-state noise and/or DC offsets appearing in the RMS energy estimates. To date evaluation of this feature has been only anecdotal. Patient MH reported an immediate improvement in clarity and naturalness of perceived speech in both quiet and noise on first application of this algorithm. She now reports only minor changes in speech quality and no changes in loudness when workplace-equivalent background noises are turned on and off in the laboratory. In addition, at her work place she had often reported hearing a brief "echo" at the end of a speech signal. This percept arose due to high level background noise mapping through her processor after speech had stopped but before the squelch delay timer had expired (section II.D.4). With the noise subtraction algorithm running, this echo effect is eliminated.

For the present portable speech processor design, two improvements are being implemented for noise subtraction. One is to perform the noise subtraction on a band by band basis as data are read from the NEC speech analysis chip before combining data to approximate the logarithmically spaced filters. The other is to compute a running average of the pure noise signal energy estimates, instead of using simply the last noise estimate obtained while the squelch was on. This would help eliminate short term and impulsive perturbations in the background noise estimates.

Formal studies of the noise subtraction algorithm using controlled noise conditions and standard speech reception tests are planned. The results will be reported in future quarterly reports.

### c. Detection of unvoiced speech energy

During each frame period while reading new spectral data from the NEC chip, a test is performed to determine whether unvoiced speech energy is present. This test requires computation of the ratio of high frequency energy (1560 - 5756 Hz) to low band energy (340 - 732 Hz) and comparison of the result to a preset threshold level. If the computed ratio exceeds the reference level, unvoiced speech sounds are considered to be present thus triggering either maximum rate or jittered rate presentation of the round-robin output frames. Since the round-robin output frame codes the current spectral information onto the output channels, abrupt increases in the output frame rate simply accentuate the high frequency spectral information associated with the unvoiced components of speech. Brief maximum rate bursts are often perceived as noise-like as well, which is consistent with the classical vocoder technique of injecting noise instead of voicing energy during unvoiced speech intervals.

High band energy is computed by combining NEC band energies 8 - 16. Low band energy is computed by combining NEC band energies 2 and 3. The ratio test is implemented by a look-up table which is precomputed assuming a fixed threshold ratio. In using the table, the high band energy value is used as the table index. The corresponding table entry is the low band energy value required to achieve the threshold ratio. The current low band energy value is compared to the value returned from the table. If the current low band energy is less, then unvoiced speech energy is detected and a flag is set. The actual stimulus output due to the detection of the unvoiced speech energy does not begin until (1) the current round-robin output frame is

completed, (2) the latest round-robin output information has been loaded into the internal RAM output buffer, and (3) the Stimulus Output Interrupt Task has recognized the unvoiced speech energy flag in that order.

The ratio value presently used for unvoiced speech energy detection is 1.0, given that the NEC gain block equalization is enabled. Of course a ratio of 1.0 makes the ratio comparison trivial, not requiring the look-up table operation described above. The ratio test originally was designed for speed, to avoid a time consuming division procedure and yet provide for fine resolution of threshold ratio values. The present ratio value of 1.0 provides for unvoiced speech energy detection for a male speaker in the absence of background noise. Further ratio evaluation is needed for female and young speakers, as well as for noisy conditions.

### 3. Stimulus train generation

Generation of the output stimulus train is a two stage process. First is the calculation of the stimulus features of the round-robin cycles, based on spectral information from the NEC chip during the current frame period. Second is the utilization of the round-robin frame information to generate the output stimulus sequence on the appropriate output channels.

To facilitate discussion, the characteristics of the round-robin stimulus frame are described now in greater detail. Figure 8 shows a single round-robin frame cycle for an interleaved-pulses processor. This particular round-robin frame is one that typically might be used for a patient who has good neuronal survival, manifested by low electrical thresholds and low channel interactions. Charge-balanced, biphasic pulses are used for each channel. In this case the biphasic pulses are short duration (200 usec/phase) and are temporally staggered across channels with short delays (100 usec interpulse time). In

FRAME PERIOD OUTPUT for  
GOOD SURVIVAL PATIENT  
(low thresholds, low interactions)

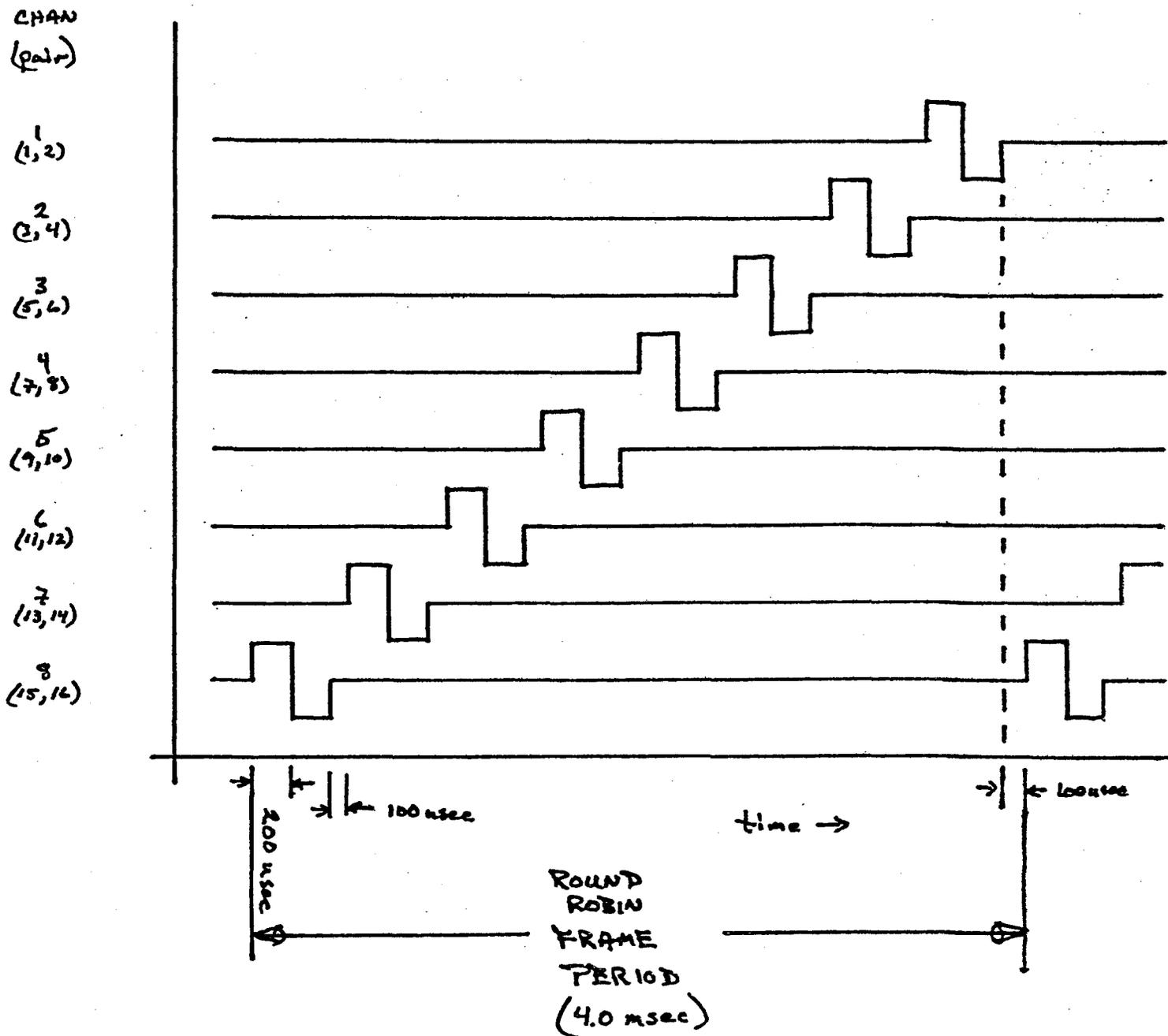
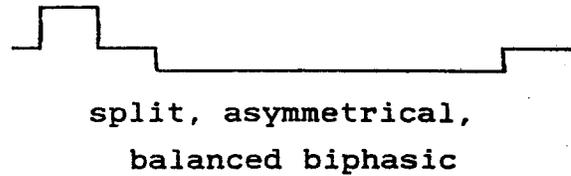
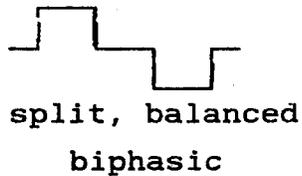
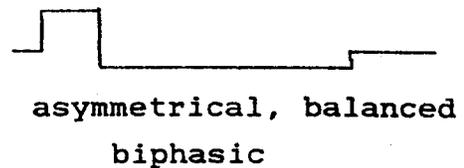
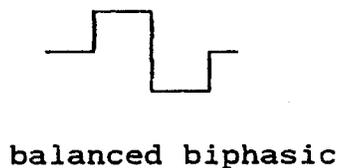


Figure 8. Output timing sequence for one round-robin frame cycle using short duration, biphasic pulses on all eight channels.

general, the pulse durations are chosen to be as short as possible yet maintain a stimulus amplitude within the operating limits of the output driver stage that can produce a MCL level percept with a 100 Hz rate, 300 msec duration pulse train. Pulse durations may differ between channels. Timing delays between pulses delivered across channels is a function of the temporal interaction time constants of the channels. Optimal times are 100 usec but can range out to 1.0 msec in poor cases with very interactive channels. As seen in Figure 8, eight channels are stimulated in a base-to-apex order (electrode pair [15,16] is the most basal bipolar pair with pair [1,2] being the most apical pair). The round-robin frame shown in Figure 8 describes a relatively simple set of stimulus characteristics that would be used with an uncomplicated, good neuronal survival patient.

The round-robin frame characteristics can change significantly for poor survival patients. These patients typically have high electrical thresholds, large channel interactions and often experience pain in conjunction with their auditory percepts. Generally high electrical thresholds force selection of long duration stimulus pulses to obtain the desired MCL level percepts. Large channel interactions lead to selection of long interpulse intervals to obtain release from interactions. Finally, in order to avoid painful stimulation and to further reduce channel interactions, manipulation of pulse wave shapes and polarities often is required.

Candidate waveshapes must maintain charge balancing. The following pulse types are prime candidates for use in interleaved-pulses processors:



Not shown are the reversed polarity conditions of each pulse shape. Wave shape and timing selection are made on a channel-by-channel basis and thus may differ among channels. No effort is made at this point to discuss the physiological basis for selecting a particular stimulus shape or timing. The purpose here is to show the diversity of the stimulus characteristics that may be needed among patients using interleaved-pulses processors.

To further complicate the design of a round-robin frame is the experimental observation that interleaved-pulses processors produce better constant confusion matrix scores if the total round-robin frame duration is held to 4 - 5 msec. This is achieved easily for the good survival patient; however, it is not achievable directly for the poor survival patient whose stimuli must be generally of longer duration. To accommodate this requirement the round-robin frame design is further modified in two ways to minimize the total frame duration.

The first method is to re-evaluate interaction characteristics once channel pulse shapes have been selected based on MCL measures and freedom from painful stimulation. Experience has shown that with selection of asymmetrical, balanced biphasic pulses, channel interactions may be dramatically reduced. This may make it possible to present stimulation on adjacent channels simultaneously, thus reducing

### Stimulus Output Interrupt Task

priority 1 interrupt driven by timer 0  
(100 usec service interval, 10 kHz)

- if within round-robin cycle, continue with round-robin cycle service
- if SQUELCH is ON, then EXIT with no output
- if internal RAM data buffer is being loaded, then EXIT with no output
- if UNVOICED speech energy flag is set, then check if it's time to start a new round-robin cycle.  
If not time, EXIT with no output.
- if PITCH PULSE FLAG set, then clear PITCH PULSE FLAG and reset begin a round-robin cycle immediately
- otherwise, EXIT with no output.

### Round-robin cycle service

- decrement timer counter
- if it is time for next event then  
write DAC value to appropriate channel  
and load timer counter with counts to  
next event in round-robin cycle.
- if it is end of round-robin cycle, then reset for new start  
otherwise, EXIT without change in output conditions.

Figure 9. Stimulus Output Interrupt Task Organization

end of the round-robin frame, stimulus outputs are stopped, unless unvoiced components of speech have been detected. If so, the interval counter is loaded with zero to initiate maximum-rate stimulation or with a random table value to initiate jittered-rate stimulation depending upon the processor design.

#### 4. Squelch and automatic gain control

Squelch and automatic gain control (AGC) processing are handled as a subtask of the NEC Service Routine because of the convenience of synchronizing AGC forward path gain changes with frame periods (8 msec) of the NEC speech analysis chip. Squelch and AGC processing are discussed separately since they are functionally independent of the spectral analysis functions of the NEC chip.

Squelch and AGC service tasks are executed once every 8 msec. These tasks essentially poll the status of the hardware flags, described in sections II.C.3 (squelch) and II.C.4 (AGC) and then reset them. The specific squelch and AGC service tasks are discussed separately.

Squelch processing begins with polling of the squelch hardware flag bit. This bit is mapped into the microprocessor via port 1 and is directly bit addressed by the software. If the flag is set, then an above-threshold signal is present to be processed. A software squelch flag, SQLON, is cleared. This automatically turns on the green LED driver indicating to the patient that a signal has been detected. In addition, a squelch delay counter is activated which forces the squelch function to remain off for about 500 msec after the signal has ceased to exceed threshold. This prevents the squelch from turning on between words during running speech. From this point, AGC processing is begun. If the squelch hardware flag is not set when polled and the squelch function is already activated, then squelch processing is exited. If the squelch hardware flag is not set when polled and the squelch function is off, then the

squelch delay counter is decremented and checked for zero. If it is zero, the SQLON is set, activating the squelch function; otherwise, squelch processing is exited. In all cases, exiting of squelch processing includes issuing a reset pulse which clears both the squelch and AGC hardware bit flags.

The AGC functions in a conventional manner with fast attack and slow release times. The forward path AGC gain control element is the programmable gain block of the NEC chip. This stage provides a range of 45 dB gain in increments of 1.5 dB.

AGC processing occurs only if the squelch mode is off and begins with polling of the AGC hardware flag which is software addressable via port 1. If the hardware flag is set, then the average RMS energy at the output of the gain block exceeds the reference level set by the VOLUME control. The processing enters the attack mode, decreasing gain at a rate of 6 dB/16 msec or 3 dB/8 msec or two 1.5 dB decrements/one 8 msec frame period. Gain changes are made by modifying a variable, GAIN, always checking to verify that GAIN stays within the acceptable 45 dB range of the NEC chip. Every time a gain decrease is made a counter/timer variable is initialized for timing the gain increments used in the release phase.

Upon polling, if the AGC hardware flag is not set, then the RMS energy at the output of the gain block is lower than the VOLUME control reference level. AGC processing then enters the release mode, increasing gain at a rate of 6 dB/320 msec or 1.5 dB/80 msec or 1.5 dB/ten 8 msec frame periods. Since the minimum gain step is 1.5 dB, long time constants must be spread step-wise across a number of frame periods. As mentioned previously, a timer/counter is initialized during attack phases and decremented once each 8 msec frame period. If the counter had reached zero, it is initialized again and a 1.5 dB gain increment is made. If the counter has not reached zero, then no gain change is made; however, the release phase processing continues.

Once the GAIN value has been chosen for the current frame period, then the gain stage of the NEC chip is adjusted. The final step in AGC processing is issuing a reset pulse which clears both the squelch and AGC hardware bit flags.

## 5. Pitch extraction

The extraction of voice pitch is a basic problem for all speech processors that explicitly encode voicing. Although useful solutions to this problem have been demonstrated, many of the methods employed for pitch extraction are too complex for analysis of speech in portable real-time devices. Included among these methods are (1) identification of high-frequency peaks in cepstrum representations of speech, (2) identification of periodic peaks in the autocorrelation function of speech that has undergone "spectral flattening," and (3) identification of pronounced discontinuities in the residual error after linear predictive analysis of speech signals.

One attractive variation of the autocorrelation method is the Average-Magnitude Difference-Function (AMDF) which can be implemented in real time (Ross, Shaffer, Cohen, Freudberg and Manley, 1974; Un and Yang, 1977). This group has previously presented the design of a real-time, AMDF-based, pitch extractor running on a 80C31 microprocessor (QPR 6, NIH project N01-NS-2356). The AMDF algorithm is a robust approach and offers good performance in poor signal to noise ratio environments (Paliwal, 1983).

For the present design, however, an alternative approach using an analog, peak-picking strategy was taken. This approach was favored initially because it has performed well as the pitch extraction method used in our Phase I bench processor. It was also felt that the the analog approach could be implemented in this new portable design more quickly than the AMDF approach. The analog approach also would conserve the 80C31 microprocessor computation resources for the needs of the post-processor until

more experience has been gained with the interleaved-pulses strategy itself.

The analog pitch extractor relies on a relatively simple method of nonlinear processing to accentuate and then detect recurring peaks in the waveforms of speech. In any periodic waveform there exists, by definition, one highest peak that is repeated in each period. If the peak is sufficiently large compared to other peaks in the period, it is possible to accentuate and then detect this peak with nonlinear shaping and analysis of the speech signal (Dolansky, 1955; Filip, 1969; and Gruenz and Schott, 1949). The circuit used for nonlinear processing of speech is illustrated in highly schematic form in Figure 11a. Waveforms present in two stages of nonlinear processing are shown in Figure 11b. for typical inputs of voiced speech. As can be appreciated from Figures 1a1 and 11b, an input signal will charge capacitor  $C_1$  when the diode of the input circuit is forward biased. When the magnitude of the input signal falls below the voltage present across the capacitor, the diode is reverse biased and ceases to conduct the input signal to the capacitor. During this period the capacitor will discharge through the low impedance path of  $R_1$ . This exponential decay of charge is shown in Figure 11b as the dashed lines that are superimposed on the input waveforms (panels a and c), and as the outputs of the first and second "detectors" (panels b and d). The process of charging and discharging capacitor  $C_1$  is repeated every time the voltage at the input exceeds the voltage across the capacitor. Thus, for the input waveforms shown, there are two charge-discharge cycles per period for "low-pitch" waves where a strong second harmonic is present, and one charge-discharge cycle per period for "high-pitch" waves where the second harmonic is relatively small.

Additional processing of the signal is performed by the differentiating circuit of  $C_2$  and  $R_2$ . This circuit emphasizes the peaks extracted by the input circuit and removes the dc component from the signal across  $C_1$ . Typical results of wave

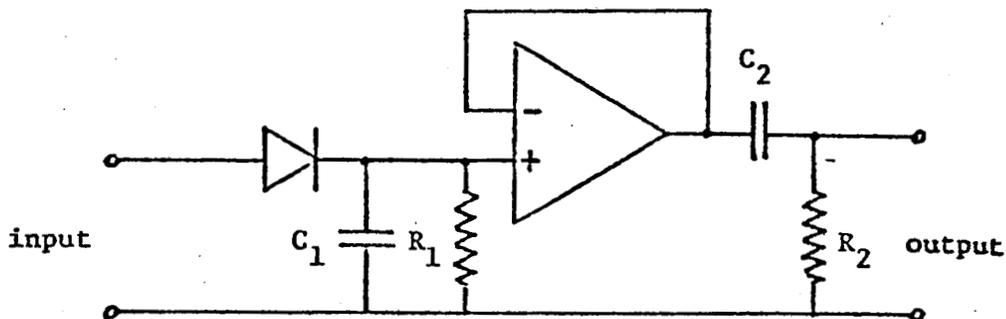


Figure 11a. Circuit used for nonlinear waveshaping.

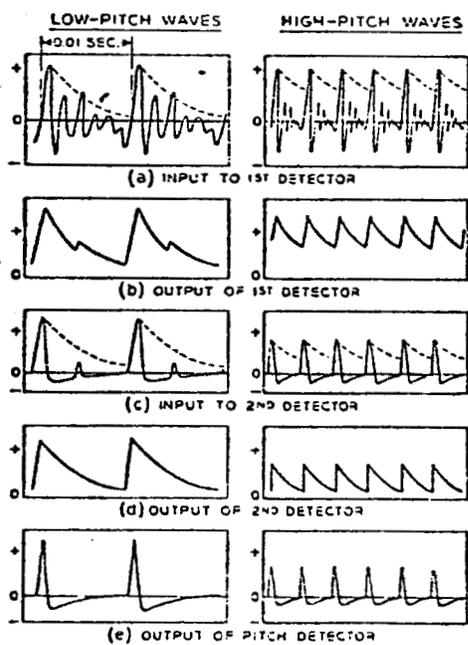


Figure 11b. Steps in the detection of periodic peaks in speech signals (taken from Gruenz and Schott, 1949).

shaping by  $C_2$  and  $R_2$  are shown in Figure 11b as the input to the second detector (panel c) and as the output of the pitch detector (panel e). The unity-gain amplifier in Figure 11a provides isolation between input and output circuits.

Two problems associated with this general method of pitch extraction are that "major" peaks in the quasi-periodic waveforms of voiced speech are rarely equal in height and that "minor" peaks often take on major proportions. False indications of voice pitch can be minimized, however, by careful selection of the time constants for the integrator and differentiator circuits, by cascading pitch extractors as shown in Figure 11b, and by analyzing both the positive and negative halves of the speech wave for periodicity detection (Filip, 1969). The optimum time constant for the integrator is a function of the expected range of voice pitch and the relative magnitudes of major and minor peaks in the speech waveform. If the time constant is too great, the pitch extractor will "skip" major peaks that follow peaks of slightly greater amplitude, and if the time constant is too small, the extractor will fail to discriminate major from minor peaks. Experiments have shown that the optimum time constant for analysis of typical speech waveforms is in the range of 4.5 to 5.0 ms.

The optimum time constant for the differentiator circuit is a function of the maximum slope of the voltage waveform across  $C_1$ , noise content of the input waveform, and available gain to offset the attenuation encountered in the process of differentiation. Because the maximum slope and noise content of the signal presented to the differentiator are different in succeeding stages of pitch extraction, the optimum time constant is different for each stage. In general, the optimum time constant for the differentiator is less than the optimum time constant for the integrator.

A pitch extractor using this approach has been included in the laboratory model of a speech-analyzing lip-reading aid for

the profoundly deaf (Cornett, Beadles and Wilson, 1978).

The performance of the pitch extractor has been evaluated using both sinusoidal and speech inputs. The dynamic range of the intensities over which the instrument will reliably extract the fundamental frequencies of sinusoids is a maximum of 26 dB at 65 Hz. False indications of pitch doubling and pitch halving are occasionally found during rapid transitions in voiced speech waveforms. The accuracy and dynamic range of the pitch extractor can be improved by (1) squaring or cubing the input signal (Sondhi, 1968), (2) center clipping the input signal (Sondhi, 1968), (3) adding an automatic gain control, and (4) processing the output signal with error logic. All four of these enhancements have been implemented in the portable speech processor.

Implementation of this analog peak-picking strategy has been done in the portable speech processor by simulating the strategy mathematically in software. As described previously in section II.C.5, the speech signal is low-pass filtered at 400 Hz and sampled at a 2 kHz rate. The signal is then center clipped and squared (Sondhi, 1968). Then, two-stage peak picking, as described above, is applied to the positive and negative signal peaks separately. Positive and negative peak information is then combined and further processed to eliminate frequency doubling. Whenever a pitch pulse (positive or negative) is identified a pitch pulse flag is set and a time-out interval of 2 msec is started and the pitch pulse magnitude is saved. If a second pulse is detected and the time-out interval is in effect, the magnitude of the pulse is compared to the previous pulse magnitude. If the new pulse is larger, then the time-out interval is reset for a full 2 msec again. If the new pulse is smaller, its occurrence is ignored. Once the time-out period expires, the next identified pulse (positive or negative) starts the process over. This post processing forces the pitch extractor to lock temporally to the largest pitch pulse, regardless of polarity. Figure 12 shows a flow diagram of the

## Pitch Extraction Interrupt Task

priority 2 interrupt driven by timer 1  
(500 usec service interval, 2 kHz)

- clock in serial input from ADC in inverse bit order (22 usec)
- do table look up using inverse bit order as index  
return value is the result after the following operations:
  - (a) data bits are placed in normal order
  - (b) data is center-clipped
  - (c) center-clipped result is then squared
- determine polarity of center-clipped and squared signal.
- for positive data
  - compare data to timer/counter 1 (PTC1)
    - if data > PTC1 then set PTC1 = data
    - if data < PTC1 then using table lookup  
decrement PTC1 as if its value  
were decaying exponentially with  
a time constant of 5.0 msec
  - compute  $d(\text{PTC1})/dt$  by subtracting previous TC1 value.  
force negative values to zero, keep only positive values.
  - compare  $d(\text{PTC1})/dt$  to timer/counter 2 (PTC2)
    - if  $d(\text{PTC1})/dt > \text{PTC2}$  then set  $\text{PTC2} = d(\text{PTC1})/dt$
    - if  $d(\text{PTC1})/dt < \text{PTC2}$  then using table lookup  
decrement PTC2 as if its value  
were decaying exponentially with a  
time constant of 2.5 msec.
  - compute  $d(\text{PTC2})/dt$  by subtracting previous  $d(\text{PTC2})/dt$  value.  
force negative values to zero, keep only positive values.
  - decrement NTC1 and NTC2 as if their values were  
decaying exponentially with time constant of  
5.0 msec and 2.5 msec, respectively.
- for negative data
  - do same as for positive data except using  
timer/counters NTC1 and NTC2 instead of PTC1 and PTC2
  - if SQUELCH is on then just EXIT
  - compare  $d(\text{PTC2})/dt$  and  $d(\text{NTC2})/dt$  to  
a variable reference level PREF
    - if  $d(\text{PTC2})/dt > \text{PREF}$ , then  $\text{PREF} = d(\text{PTC2})/dt$   
and branch to time out test

Figure 12. Pitch Extraction Interrupt Task  
(continued to next page)

```
if d(NTC2)/dt > PREF, then PREF = d(NTC2)/dt
    and branch to TIMEOUT TEST

    otherwise,
        decrement TIMEOUT counter

- if TIMEOUT period is over, then reset PREF to its original
  fixed threshold value

- EXIT

- TIMEOUT TEST

    if TIMEOUT period in effect, then
        reset time out counter for full period
        and EXIT

    otherwise, SET PITCH PULSE FLAG
        and set time out counter for full period
        and EXIT
```

Note: The PITCH PULSE FLAG is reset by the Stimulus Output Interrupt Task once the presence of the pitch pulse has been recognized and its associated round-robin frame initiated.

Figure 12. Pitch Extraction Interrupt Task

software logic used in the portable processor.

## 6. Mode selection

Mode switch settings are read by the background processing task. Four modes are possible. A diode switching matrix codes the four switch positions as a two-bit binary code on the microprocessor port 1. The software reads these bits and compares the new switch data to the previous switch data. If no change has occurred, the background task continues its normal looping. If a change has occurred, then flags are set to indicate the new mode condition, any immediate processing changes are made (i.e., changing parameters or jump commands, etc.) and then normal background processing resumes.

The mode switch option is a valuable feature to have in a microprocessor-based processing system. We utilize this option to provide the patient with a range of operating modes that she may select. In this manner, the patient may directly compare two or more processing strategies in a particular listening situation. Ultimately, we envision that a patient may be fitted with perhaps three distinct speech processing strategies all available in the same instrument -- one optimized for best performance in quiet, one optimized for best performance in noise with lip-reading (e.g. a two-channel Breeuwer/Plomp strategy) and one optimized for conveying the most salient features of that patient's favorite music.

## E. Stimulation Output

The output driver stages of the portable speech processor are transformer coupled voltage sources capable of + and - 10 volts and 20 kHz bandwidth. Six independent channels are presently provided. A driver for a single channel consists of a memory-mapped, latched CMOS DAC (one stage of a quad DAC) feeding a two stage operational amplifier driver that is transformer-coupled to the electrodes. The DAC operates with a unipolar binary code and uses a DC reference source common to all channels. The DC reference level is adjustable with an internal potentiometer and provides a master gain control across all six channels. The electrodes are driven in a bipolar manner and are capacitively-coupled to the transformer output coils. The transformers are the bandwidth limiting elements of the output driver stages with a typical frequency response of 300 - 20 kHz. The 20 kHz upper cutoff provides good rise time characteristics for the pulsatile outputs; however, the 300 Hz lower cutoff contributes substantial decay distortion to the pulse wave forms.

As configured, this output driver stage is the least desirable aspect of the present system. The initial choice to use a voltage-controlled as opposed to a current-controlled output driver was made to expedite the development of the portable processor. The voltage-controlled stages did not require power supply isolation between channels because of the isolation provided by transformer coupling. In contrast, current-controlled drivers require full supply isolation (including grounding) between channels to insure control of the current through the electrodes. Such a design is technically possible, but could not be implemented quickly and yet maintain power consumption requirements consistent with a battery-operated, portable unit. Therefore, the decision was made to begin with the transformer-coupled driver stages and later retrofit the more desirable current-controlled outputs.

## F. Power consumption

As presently described the portable interleaved-pulses processor consumes approximately 600 mWatts of power. This roughly breaks down to the following power budget:

NEC speech analysis circuit	250 mWatts;
Output driver board	200 mWatts;
Front-end hardware processing and microprocessor (8 MHz)	150 mWatts.

A single battery pack provides approximately 4500 mWatt-hrs (5 volts x 900 mAmp-hr) of energy, thus allowing six to seven hours of operation on a single battery pack. These figures, although higher than ultimately desired, are consistent with our original design criteria for an initial, portable speech processor that provides full flexibility in implementing interleaved-pulses processing strategies.

Several options exist for reducing power consumption of the present design. One is to replace the switched-capacitor filter bank of the NEC speech analysis chip with a linear, active-filter, band-pass filter bank. Of course, additional electronics must be added for RMS calculation, AGC function and microprocessor interfacing; however, preliminary calculations suggest that these functions combined would require 40 to 50 mWatts at most. The output driver board power consumption can be easily reduced with better transformer selection and reducing operating voltages from + and - 12 volts to 0 to 3.5 or 5 volts. A general redesign of this driver board is underway in any case to provide current-controlled sources. Front-end processing hardware and microprocessor power may be reduced by dropping operating voltages, reducing the microprocessor clock frequency and using the idle down operating feature of the microprocessor in conjunction with the squelch circuit.

## G. Packaging

This section describes the physical package configuration of the portable unit. The package was designed for the specific needs of patient MH. MH presently has a percutaneous cable connection that exits just above her left ear (her right ear is implanted). In the interest of safety and convenience, the processor is packaged in a small plastic case that is placed within a custom designed carrying pouch and strap. The strap of the carrying pouch is worn across the right shoulder and passes across the body to the left hip where the processor pouch is located. Figure 13 shows a patient wearing the processor while making a front control panel adjustment.

The strap is a hollow envelope and contains both the microphone cable and the six-channel (twelve-lead) output cable. As seen in Figure 13, the microphone is located about two-thirds the way up the strap, passing through a slit in the strap itself. Just above the microphone, at about clavicle level is the output cable connector. MH passes her percutaneous cable behind her neck to this output connector. This connector is hidden beneath the strap against the patient's clothing. When connected, the connections lie flat just above the right clavicle, and lie unseen beneath the carrying strap.

Figure 14 is another view of the carrying case showing the access for the rechargeable battery. Battery swaps are made by folding down the large flap that folds over the side of the processor and opening the two thin end flaps. All flaps are held in place using plastic hook-and-loop fastening material. The battery drops out of the bottom of the case, still attached to the power cable. As seen on the diagram drawn on the large side flap, the power cable unplugs from the battery pack. Battery swaps are easily made without removing the carrying pouch and strap from the body. Such swaps require about a minute to complete.

Percutaneous  
cable connector

Microphone

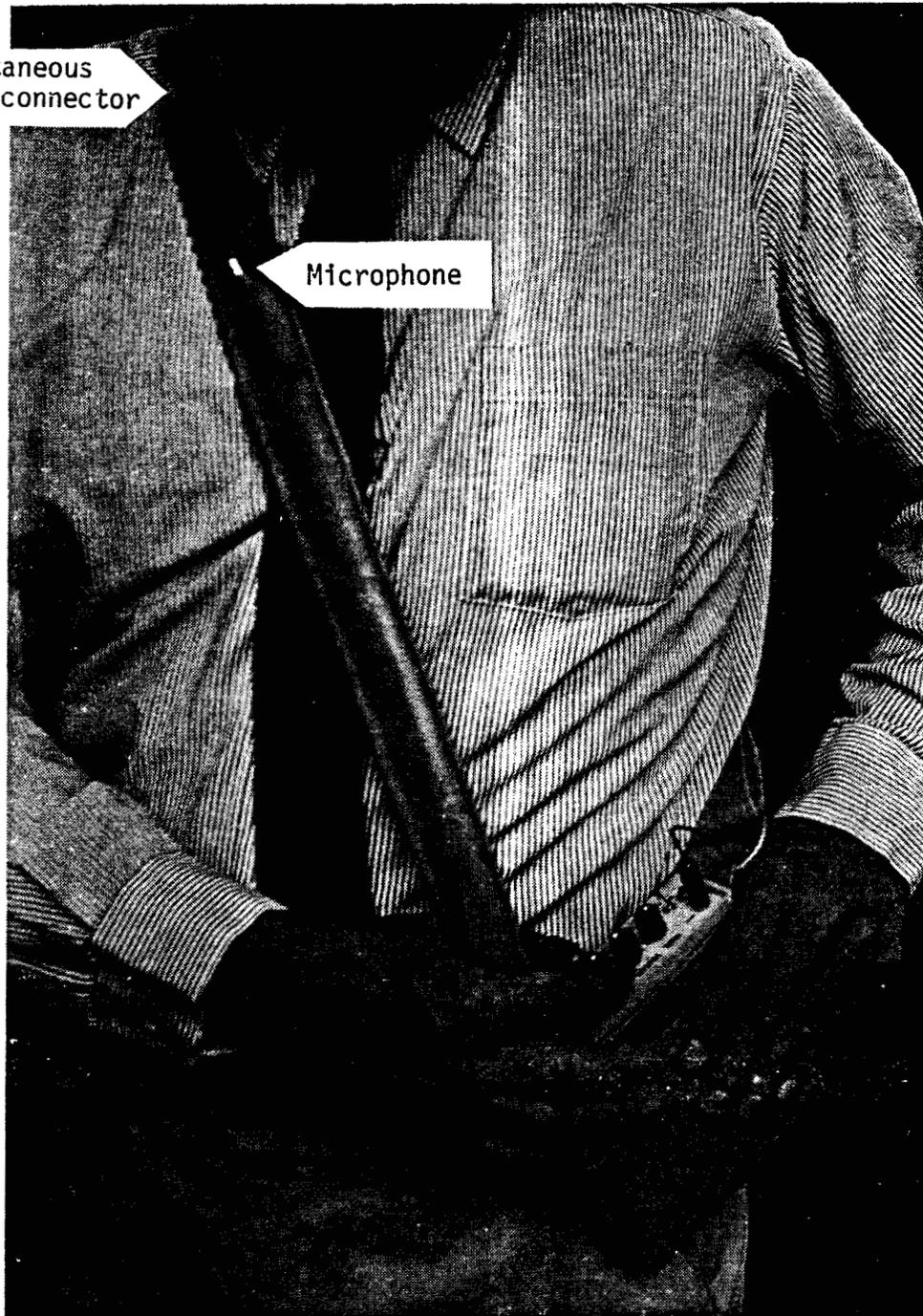


Figure 13. View showing patient wearing the portable speech processor while making a front control panel adjustment.

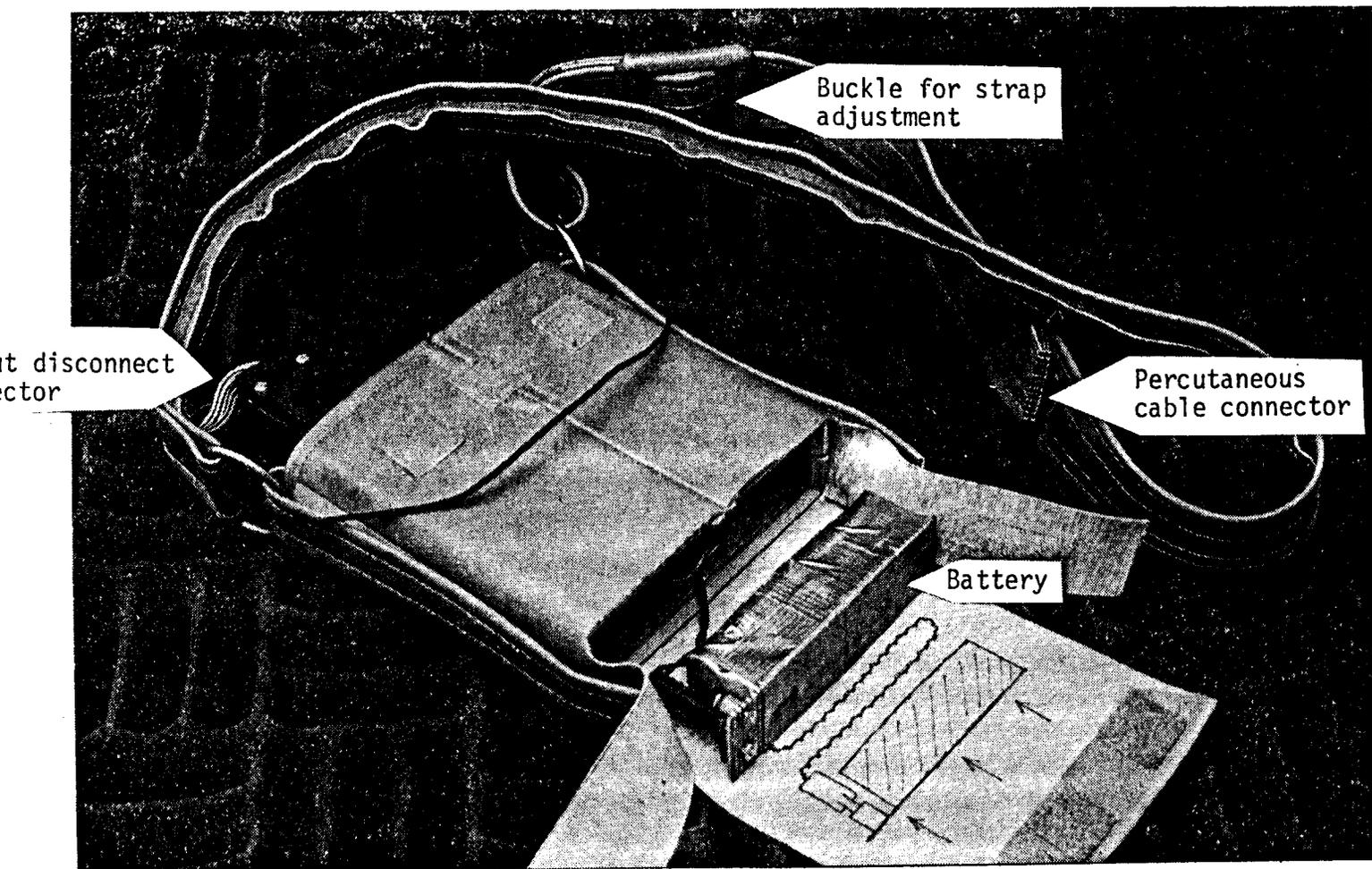


Figure 14. View showing wearable speech processor with access flaps open and battery pack partially removed.

Also seen in Figure 14 are the microphone cable and output cable exiting from the hollow shoulder strap where it attaches to the carrying case. The shoulder-level output connector that is normally concealed beneath the strap during use, may also be seen. The buckle seen in the top of the figure is worn across the back and provides for adjustment of the shoulder strap length.

Figure 15 shows the front view of the processor in its carrying pouch. This is the basic view of the processor that the patient has while wearing the unit. In the center of the front panel are three knobs that are, from left to right, the mode switch, the squelch adjustment and the volume adjustment. To the right of the volume knob is the stimulus output connector. The connector plug for this socket is seen in Figure 14. This large plug provides an easily accessed and quickly removed safety disconnect in the event the patient needs to quickly disable the unit. In addition, this plug is the last to be connected during start up and thus serves to minimize the patient's exposure to start-up transients. Beneath the output connector are two LED indicators. One LED indicator (GREEN) indicates when the squelch circuit is disabled, that is to say that it is on when signals are being processed and stimuli delivered to the patient. This visual LED signal is helpful to the patient in adjusting the squelch circuit threshold. Typically the squelch is adjusted to discriminate the presence or absence of the patient's own voice. The green indicator is wired in series with a normally open push button switch to conserve battery life once the squelch control has been adjusted. The other LED indicator (RED) flashes when the battery terminal voltage had dropped beneath its minimum terminal voltage (4.5 V typically). On power up, both the red and green LED indicators are used to flash a characteristic pattern to affirm to the patient that the processor has passed its own self test and is working properly. At that point, the patient plugs in the electrode cable to the output connector. At the extreme left of the unit are the microphone input connector,

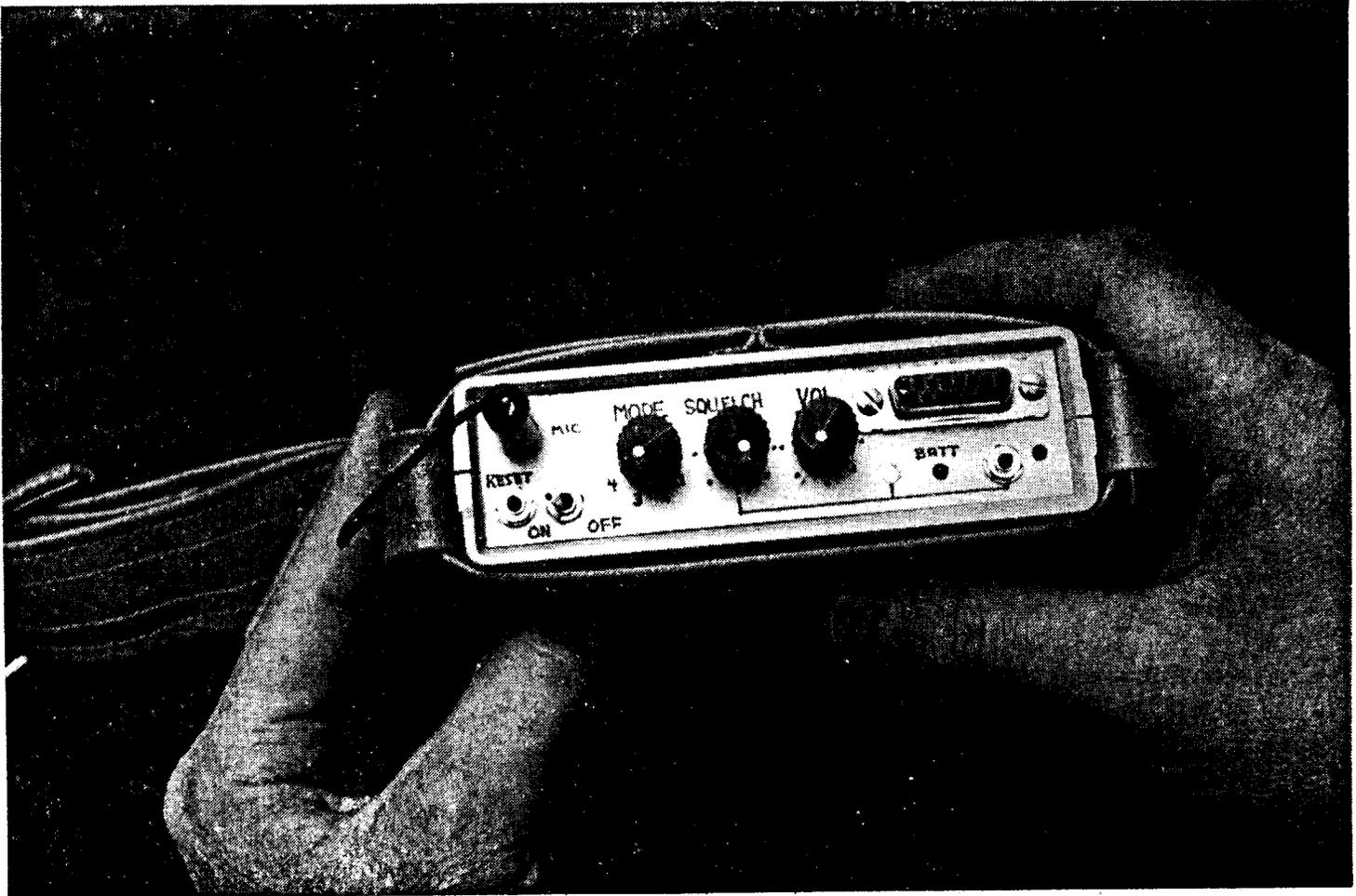


Figure 15. View showing front panel of wearable speech processor with output cable connector removed.

a master reset for the microprocessor and the main on/off switch for the battery.

Figure 16 shows the processor turned upside down and with the bottom case removed. This view shows the compact construction of the unit. Two layers of perforated boards are used, each being hand wired point-to-point. The component sides of the boards are mounted inward. The rechargeable battery pack is seen at the back end of the unit. The connector strip seen at the left front corner of the upper board carries output stimuli down to the front panel output connector. Disconnecting this connector at the edge of the board and removing the battery pack allows the processor boards to be folded open, as seen in Figure 17.

With the board folded open, the board still inside the case and closest to the front panel contains the front-end processing hardware, the microprocessor with memory, and the NEC speech analysis chip. The other board contains the DACs and driver circuitry for generating the output stimuli for six channels.

Figure 18 is a view of the battery charger that accompanies the unit. Four batteries are provided. Each battery pack contains eight AA nicad rechargeable batteries (1.25 volts at 450 mA-Hr.) wired into two parallel stacks of four batteries. Each pack provides typically 5 volts at 900 mA-Hr. Charging is at C/10 rate. Each battery pack provides approximately 6 hours of operation of the processor.

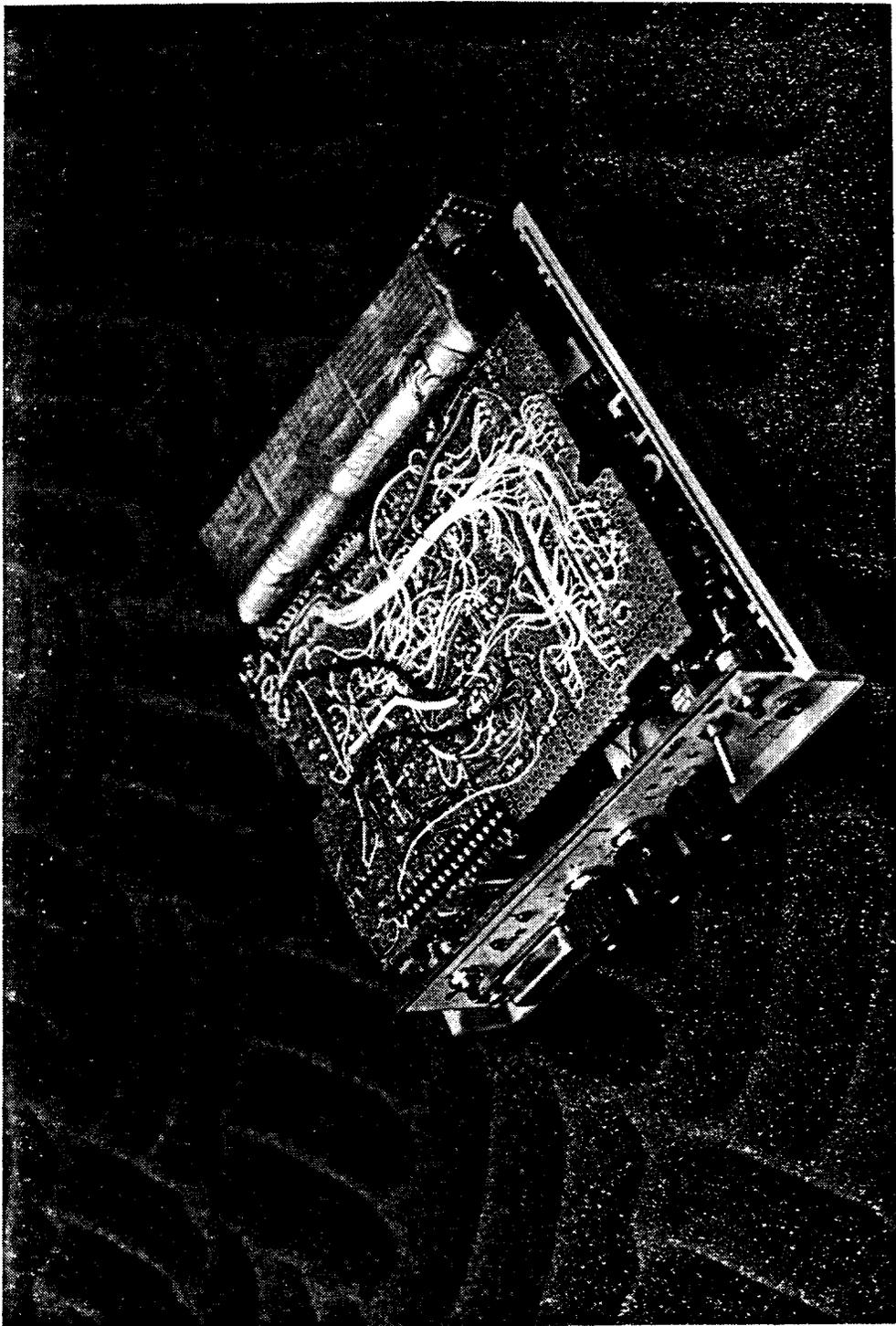


Figure 16. View showing wearable speech processor with case removed. Bottom of output driver board is seen with rechargeable battery pack at back.

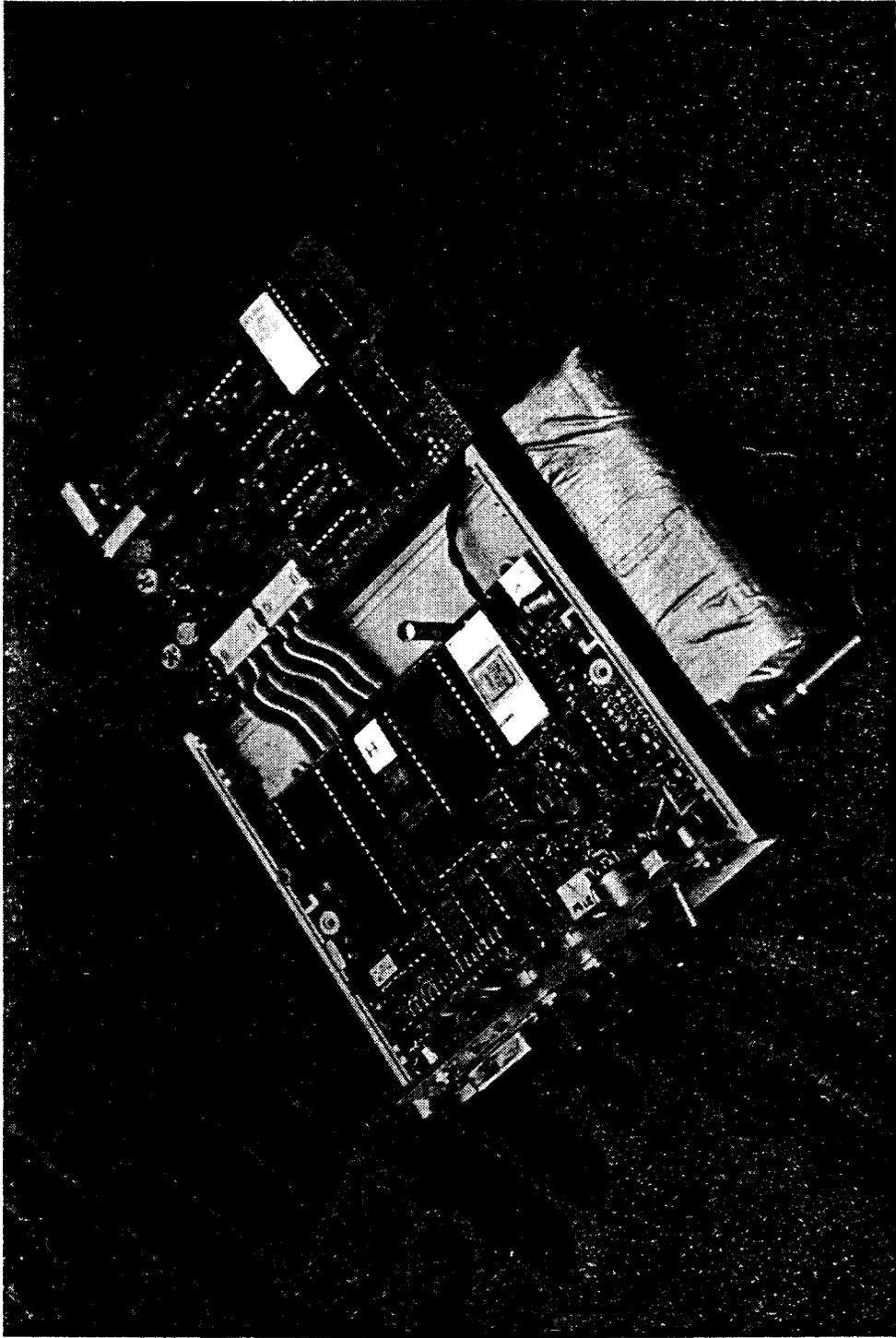


Figure 17. View showing wearable speech processor with output driver board folded to the right, exposing front-end hardware and microcontroller board.

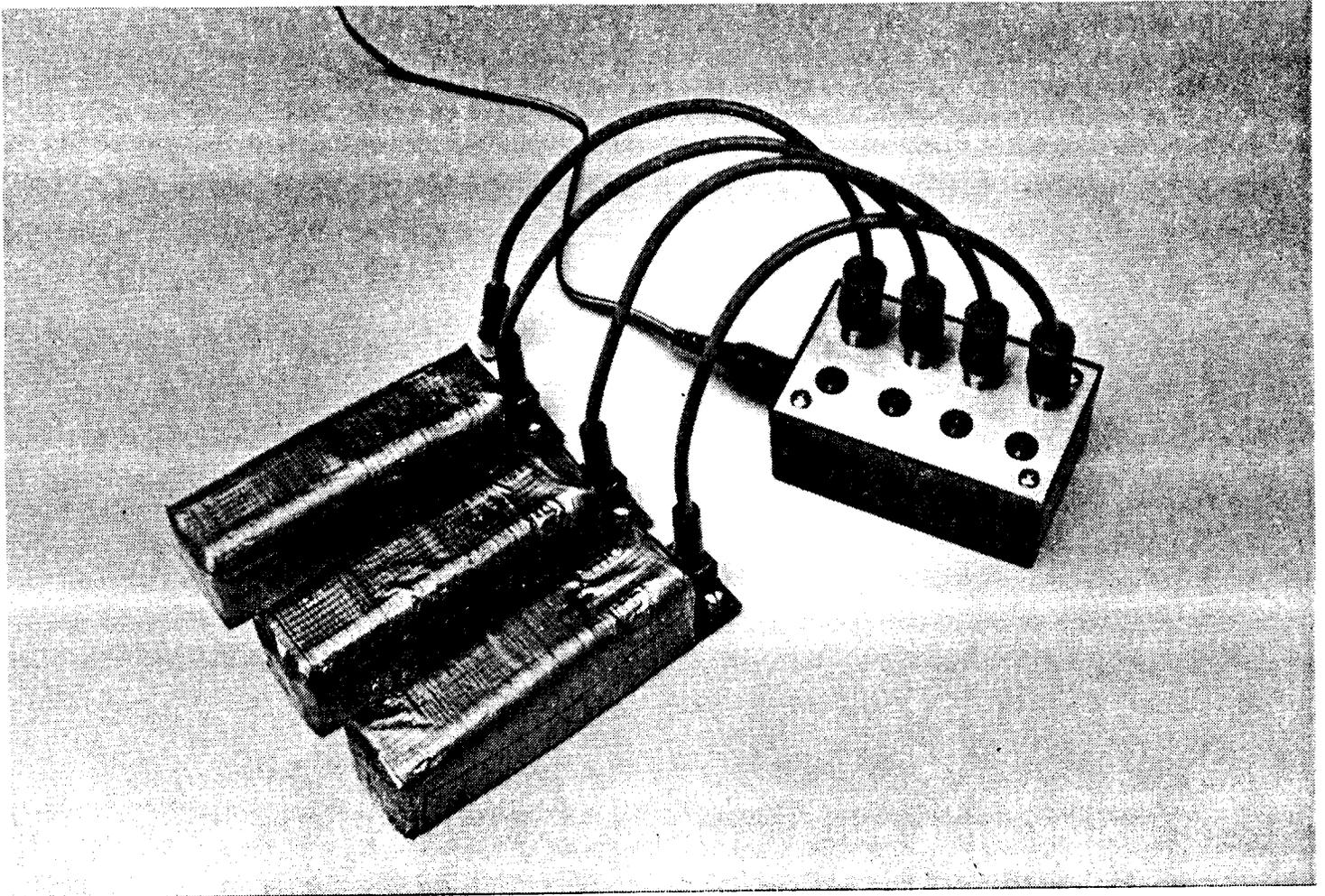


Figure 18. View showing battery charger and three battery packs. Cable at top of figure leads to a 110 V AC to 9 V DC wall plug module.

## H. Future improvements

Several significant improvements are being considered for the next design iteration. Most of these have been discussed earlier in their appropriate sections. A brief list is provided here.

1. Conversion of voltage-controlled drivers to current-controlled drivers, accompanied by reduction of power consumption.
2. Switch to alternative pitch extraction techniques that are more robust in noise. The most attractive candidate is the AMDF technique (QPR 6, NIH project NO1-NS-2356).
3. Evaluation of the Motorola MC68HC11 microprocessor to replace the Intel 80C31 processor presently used, thus reducing power consumption.
4. Replacement of the NEC speech analysis circuit with low-power, linear, active-filter bank and associated electronics for AGC, RMS calculation and microprocessor interfacing to reduce power consumption.

We also are investigating the feasibility of directly sampling the multiplexed band-pass output of the NEC filter chip. This would allow the same hardware to simulate both the present UCSF/Storz analog processor and the interleaved-pulses strategies for more direct comparison of the two approaches.

### III. PLANS FOR THE NEXT QUARTER

Our plans for the next quarter include the following:

1. Continue development of the portable speech processor, as indicated in section II. H. of this report;
2. Resume work on the "integrated field-neuron" model, as indicated in part in Appendices 3 and 4 to this report;
3. Initiate follow-up studies for selected patients in our last series of six patients fitted with the 4-channel, UCSF/Storz transcutaneous transmission system (see QPR 6 for a brief description of the initial studies);
4. Initiate collaborative studies with Bryan Pfingst of the Kresege Hearing Research Institute to extend our investigations of loudness and pitch perception with patient MH;
5. Present project results in three invited lectures at the Gordon Conference on "Implantable Auditory Prostheses" (June 29 through July 3, 1987) and in one invited lecture at the International Cochlear Implant Symposium 1987 (September 7 through September 11, 1987); and
6. Prepare two invited papers, one for a chapter in Cochlear Implants 1987, to be published by Springer-Verlag, and the other for the upcoming special issue of the Proc. IEEE on "Emerging Electromedical Systems," to be published in September, 1988.

## REFERENCES

1. Cornett, R.O., Beadles, R.L. and Wilson, B.S., "Automatic Cued Speech," Proc. Research Conference on Speech-Processing Aids for the Deaf, 1978, 19 pp.
2. Dolansky, L.O., "An Instantaneous Pitch-Period Indicator," J. Acoust. Soc. Am., 27: 67-72, 1955.
3. Filip, M., "Envelope Periodicity Detection," J. Acoust. Soc. Am., 45: 719-732, 1969.
4. Gruenz, O.O. and Schott, L.O., "Extraction and Portrayal of Pitch of Speech Sounds," J. Acoust. Soc. Am., 21: 487-495, 1949.
5. Paliwal, K.K., "Comparative Performance Evaluation of Different Pitch Estimation Methods for Noisy Speech," Acoustics Letters, 6(11): 164-166, 1983.
6. Ross, M.J., Shaffer, H.L., Cohen, A., Freudberg, R. and Manley, H.J., "Average Magnitude Differences Function Pitch Extraction," IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-22(5): 353-362, 1974.
7. Sondhi, M.M., "New Methods of Pitch Extraction," IEEE Trans. on Audio and Electroacoustics, AU-16: 262-266, 1968.
8. Un, C.K. and Yang, S.C., "A Pitch Extraction Algorithm Based of LPC Inverse Filtering and AMDF," IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-25(6): 565-572, 1977.

Appendix 1

Summary of Reporting Activity for the Period of March 26  
through June 26, 1986, NIH Contract N01-NS-5-2396

In this quarter the following manuscripts were submitted and accepted for publication:

Wilson, BS, CC Finley, JC Farmer, Jr., DT Lawson, BA Weber, RD Wolford, PD Kenan, MW White, MM Merzenich and RA Schindler: Comparative studies of speech processing strategies for cochlear implants, to be published in Laryngoscope.

Wilson, BS, CC Finley, MW White and DT Lawson: Comparisons of processing strategies for multichannel auditory prostheses, to be published in 1987 IEEE Frontiers of Engineering in Health Care.

White, MW, CC Finley and BS Wilson: Electrical stimulation model of the auditory nerve: Stochastic response characteristics, to be published in 1987 IEEE Frontiers of Engineering in Health Care.

Finley, CC, BS Wilson and MW White: A finite-element model of bipolar field patterns in the electrically stimulated cochlea - a two dimensional approximation, to be published in 1987 IEEE Frontiers of Engineering in Health Care.

The first manuscript was included as an appendix to our last quarterly report for this project, and the remaining manuscripts are included as appendices to the present report.

In addition to submission of the above papers for publication, the following presentations were made in the present reporting period:

Lawson, DT: Cochlear implants. Invited presentation in the UNC-Greensboro series of Psychology Colloquia, April 10, 1987.

Wilson, BS et al.: Comparative studies of speech processors for cochlear implants, presented at the National Meeting of the Triological Society, Denver, CO, April 27, 1987.

Finley, CC and RD Wolford: Cochlear implants. Invited lecture, Departments of Otolaryngology and Audiology, UNC School of Medicine, May 8, 1987.

Wilson, BS: The RTI/Duke cochlear implant program, presented to the Executive Committee of the Research Triangle Institute (RTI) Board of Governors, June 17, 1987.

Farmer, JC, Jr. and BS Wilson: Cochlear implantation for the profoundly deaf. Invited seminar presentation, Department of Physiology, Duke University Medical Center, June 18, 1987.

## COMPARISONS OF PROCESSING STRATEGIES FOR MULTICHANNEL AUDITORY PROSTHESES

Blake S. Wilson<sup>1,2</sup>, Charles C. Finley<sup>1</sup>, Mark W. White<sup>3</sup> and Dewey T. Lawson<sup>1</sup><sup>1</sup>Neuroscience Program, Research Triangle Institute, RTP, NC 27709.<sup>2</sup>Division of Otolaryngology, Duke Medical Center, Durham, NC 27710.<sup>3</sup>Electrical & Computer Engineering Dept, NC State U, Raleigh, NC 27695.

## ABSTRACT

We have compared a wide variety of processing strategies for auditory prostheses in studies with two patients implanted with the UCSF electrode array. Each strategy was evaluated using tests of vowel and consonant confusions. Included were the compressed-analog (CA) strategy of the present UCSF prosthesis and a group of interleaved-pulses (IP) strategies in which the amplitudes of non-simultaneous pulses code the short-time spectra of speech. For these patients, each with indications of poor nerve survival, scores were significantly higher with the IP processors. We believe an important contributor to this superior performance is the substantial "release" from channel interactions provided by non-simultaneous stimuli.

## INTRODUCTION

A fundamental problem faced by designers of cochlear implants is the generally sparse and uneven survival of neural elements in the typical deaf ear. In cases where survival of dendrites or ganglion cells is poor, relatively high levels of electrical stimulation are required to elicit auditory responses and the electric fields from different electrodes generally overlap to excite common subpopulations of neurons. This overlap in excitation fields is the problem of channel interactions, which can severely limit the performance of multichannel prostheses.

On the other hand, some deaf ears have a relatively complete survival of ganglion cells and dendrites. In such cases, low levels of electrical stimulation can be used, and the problem of channel interactions is much less severe.

In this paper we will describe a new processing strategy designed to minimize the deleterious effects of channel interactions. In addition, we will present results obtained in direct comparisons of a conventional processing strategy and the new processing strategy in tests with two implant patients. Both of these patients were implanted with the UCSF electrode array and fitted with percutaneous cables. Direct access to the implanted electrodes via the cable allowed evaluation of many alternative processing strategies.

## COMPRESSED-ANALOG PROCESSORS

A widely-applied processing strategy for multichannel implants is illustrated in Fig. 1. We call it the Compressed Analog (CA) strategy because the basic functions of the processor are first to compress the wide dynamic range of input speech signals into the narrow dynamic range of electrically-evoked hearing, and then to filter the compressed signal to parcel out frequency components in speech. Typical waveforms of such a processor are shown in the figure. The top trace in each panel is the input signal, which is the word "BOUGHT." The other waveforms in each panel are the filtered output signals for 4 channels of intracochlear stimulation. The bottom-left panel shows an expanded display of waveforms during the initial part of the vowel in BOUGHT, and the bottom-right panel shows an expanded display of waveforms during the final "T." The lower panels

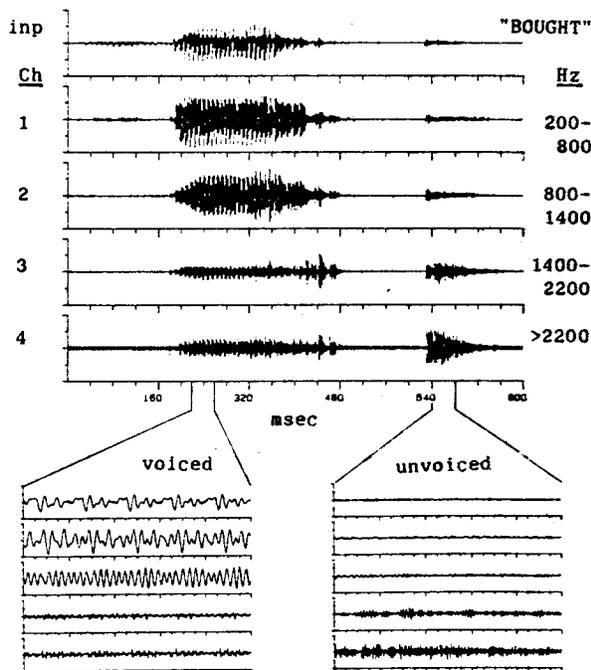


Fig. 1. Waveforms of a compressed-analog (CA) processor.

thus indicate differences in waveforms for voiced and unvoiced phonemes of speech.

In the voiced interval the relatively large outputs of channels 1 and 2 reflect the low-frequency content of the vowel, and in the unvoiced interval the relatively large outputs of channels 3 and 4 reflect the high-frequency noise content of the "T." In addition, the clear periodicity in the waveforms of channels 1 and 2 reflects the fundamental frequency of the vowel during the voiced interval, and the lack of periodicity in the outputs of all channels reflects the noise-like quality of the "T" during the unvoiced interval. These represented features are likely to be perceived to varying degrees by different implant patients. A major concern is that simultaneous stimulation with continuous waveforms can produce summation of the electric fields from the individual electrodes. This summation can exacerbate interactions between channels, particularly for patients who require high stimulation levels. Therefore, one might expect that CA processors would work best for patients with low thresholds and good isolation between channels.

#### INTERLEAVED-PULSES PROCESSORS

The problem of channel interactions is addressed in the processor of Fig. 2 with the use of non-simultaneous stimuli. The stimuli consist of brief pulses presented in sequence to the different channels. There is no overlap between the pulses so that direct summation of electric fields produced by different electrodes is avoided. The energies of frequency components in speech are represented by the amplitudes of the pulses, and distinctions between voiced and unvoiced segments of speech are represented by the timing of sequences of stimulation across the electrode array. In this particular processor stimulus sequences begin in synchrony with the fundamental frequency for voiced speech sounds and at randomly-spaced intervals for unvoiced speech sounds. The periodicity of stimulus sequences can be seen for a voiced speech sound in the lower-left panel and the randomly spaced stimulus sequences can be seen for an unvoiced speech sound in the lower right panel.

Because non-simultaneous stimuli are used in this "Interleaved-Pulses" (IP) processor, one might expect that its use could improve performance for patients with large channel interactions.

#### PATIENT LP

The first patient had a most discouraging picture of psychophysical performance. He had extremely severe channel interactions and high thresholds for stimulation of bipolar pairs in his electrode array. In addition, his case was further complicated by extremely narrow dynamic ranges and lability of thresholds and loudness levels both within and between testing sessions. In all, his psychophysical findings were consistent with very poor survival of neurons in his implanted ear.

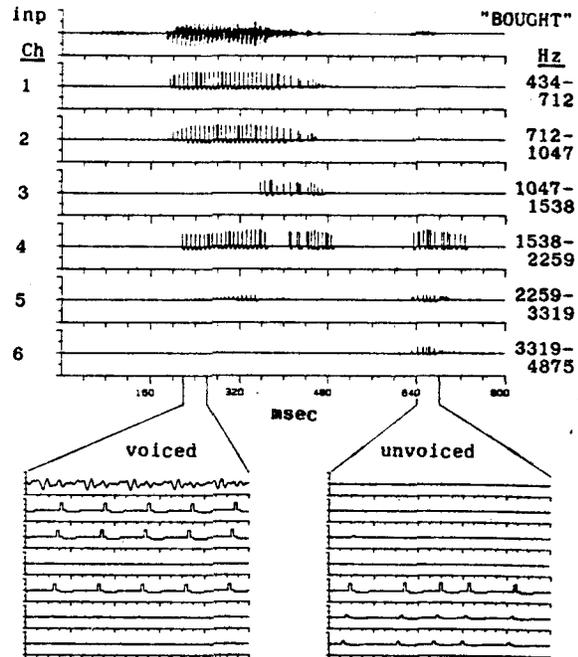


Fig. 2. Waveforms of an interleaved-pulses (IP) processor.

As might be expected, LP received no benefit from the CA processor used in the standard UCSF prosthesis. He refused to describe any of the percepts produced with this processor as speechlike, and his scores on 13 of the 14 speech tests in the Minimal Auditory Capabilities Battery did not exceed chance levels of performance.

Fortunately, the use of IP processors for this patient produced large gains in performance over the CA processor. We are pleased to report that the first application of a 6-channel, IP processor immediately moved LP into the speech mode of auditory perception. A record of LP's initial reports in listening to the outputs of this processor is shown in Table 1. Of the 11 speech tokens presented, 7 were immediately and spontaneously recognized as the correct words or syllables. In most cases where spontaneous recognition was not achieved, LP could identify

Table 1. Initial reports made by LP in listening to the outputs of an IP processor.<sup>a</sup>

Vowels:	BOOT	BOUGHT*	BOAT*	BIT*	BEET*	
Consonants:	ADA*	AKA	ANA*	ASA	ATA*	AZA

<sup>a</sup>Tokens ALA and ATHA were not presented in the initial tests with this first IP processor.

\* Indicates spontaneous recognition.

the correct class of speech sound. For example, when AZA was presented LP said that it could either be ASA or AZA, but there was no way he could tell the difference between these two. In all, the improvement over the very disappointing results obtained with the CA processor was immediate and compelling. Additional formal tests with vowel confusion matrices indicated that LP could perform at a level well above chance with a reduced 4-channel version of the IP processor. Although his performance seemed to decline when the number of channels was reduced from 6 to 4, both IP processors provided an enormous improvement over the CA processor.

#### PATIENT MH

With our second patient we were able to evaluate differences in processor performance in much greater detail. This patient also had psychophysical manifestations of poor nerve survival. The results presented in Fig. 3 show her levels of performance in formal tests of vowel and consonant recognition. The diagonally-hatched bars show her performance with lipreading, and the double-hatched bars show her performance without lipreading. The chance levels of performance are indicated by the horizontal lines in each panel. Finally, different processors are represented by different sets of bars. The characteristics of these processors are indicated in the bottom set of labels. For example, the leftmost set of bars shows the scores for a 4-channel, CA processor. The remaining sets of bars show the scores for four variations of IP processors. These variations were produced by manipulating the number of stimulation channels and by manipulating the way in which the beginnings of stimulus sequences were timed. In one variation stimulus sequences were timed to start in synchrony with the fundamental frequency for voiced speech sounds and at randomly-spaced intervals during unvoiced speech sounds. These processors had explicit coding of fundamental frequency and voice/unvoice distinctions. In the other variation stimulus sequences were timed to follow each other as rapidly as possible. These processors did not have explicit coding of voicing information. In all, the results shown in Fig. 3 allow direct comparisons of (1) a 4-channel CA processor and a 4-channel IP processor; (2) 4- and 6-channel IP processors; and (3) 4- and 6-channel IP processors with and without explicit coding of voicing information. The comparisons indicate that:

First, Performance is markedly improved when a 4-channel IP processor is used instead of a 4-channel CA processor;

Second, Scores are much higher when a 6-channel IP processor is used instead of a 4-channel IP processor; and

Third, Explicit coding of voicing information improves the performance of IP processors.

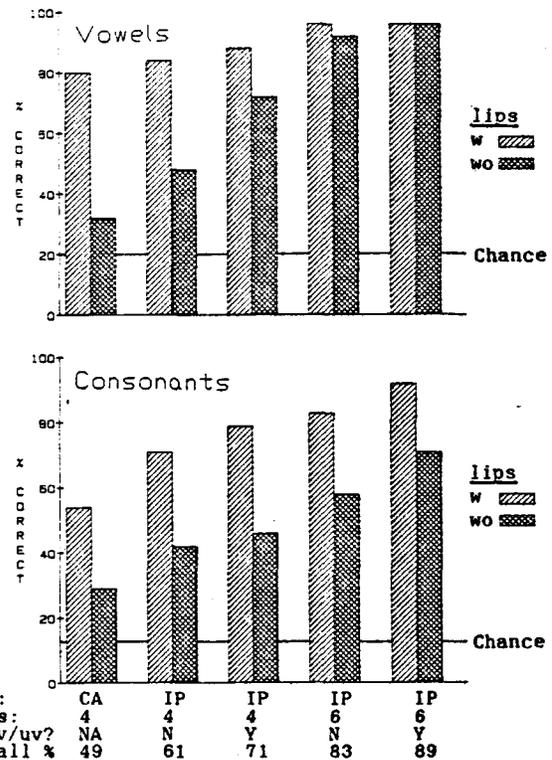


Fig. 3. Results of vowel and consonant confusion tests for patient MH.

#### CONCLUDING REMARKS

In conclusion, we observe that:

1. Different processing strategies can produce widely-different outcomes for individual implant patients;
2. Interleaved-pulses processors are far superior to other processors for at least two patients with poor nerve survival;
3. Processors other than the interleaved-pulses processors may be superior for patients with good nerve survival; and
4. Substantial gains in speech recognition can be made by selecting the best type of speech processor for each patient.

#### ACKNOWLEDGEMENTS

We thank the patients who participated in this study for their dedicated effort and pioneering spirit. We are also pleased to acknowledge the important contributions made by MM Merzenich, RA Schindler, JC Farmer, Jr., BA Weber, RD Wolford and PD Kenan. This work was supported by NIH contracts N01-NS-3-2356 and N01-NS-5-2396, from the Neural Prosthesis Program.

## ELECTRICAL STIMULATION MODEL OF THE AUDITORY NERVE: STOCHASTIC RESPONSE CHARACTERISTICS

Mark W. White<sup>1</sup>, Charles C. Finley<sup>2</sup>, Blake S. Wilson<sup>2</sup><sup>1</sup>ECE Dept, NC State University, Raleigh, NC 27695.<sup>2</sup>Neuroscience Program, Research Triangle Institute, RTP, NC 27709.

## INTRODUCTION

We are developing a neurophysiological model for electrical stimulation of the auditory nerve. Such a model will improve our ability to design electrode arrays and speech processing systems for cochlear implant recipients. In this brief report we will only consider the stochastic component of such a neural model. An understanding of the stochastic aspects of electrical excitation may be essential for accurate interpretation of patient performance with cochlear implants.

In this report we will consider an important relationship between the anatomy and the physiology of auditory nerve fibers. In particular, Verveen's stochastic model [5] allows us to predict the shape of a fiber's probability-of-firing function based on simple anatomical measurements.

## THE STOCHASTIC NEURAL MODEL

Verveen [4,5] has postulated that neural membrane noise is generated by individual ions or packets of ions as they pass through channels in the semi-permeable membrane. The flow of ions can be described as the sum of two components: one representing the steady-state or average rate of flow (i.e. the DC component of the ionic currents), and the second representing the stochastic variations around the steady-state average. In large nodes of Ranvier, with presumably large numbers of independent conduction channels, the stochastic component of current flow will be relatively small when compared to the average current flow. [ This is analogous to what occurs in standard signal averaging. As we increase the number of trials (analogous to increasing the number of conduction channels) the noise becomes less significant when compared to the average.] For very small nodes, the stochastic current becomes a very significant component of the total membrane current. This is particularly relevant because auditory nerve dendrites can be quite small.

Verveen has measured how the probability of discharge varies with stimulus intensity for a wide range of fiber diameters. For a large range of stimulus conditions, estimated discharge

probability versus stimulus amplitude functions were well fit by integrated Gaussian functions. This finding is consistent with a model where Gaussian noise is added prior to an ideal threshold detector. The idealized threshold detector "generates" an action potential whenever the voltage at its input reaches or exceeds the threshold voltage. The RMS noise voltage  $V_{noise}$  and the threshold voltage  $V_{thr}$  completely determine how the probability of firing will vary with stimulus amplitude. Normalizing  $V_{noise}$  with respect to  $V_{thr}$  provides a relative noise level measure  $R$  defined as:

$$R = V_{noise} / V_{thr}$$

Nodes with low normalized noise levels (small  $R$ ) produce steep input-output functions.

Equivalently, the steepness of the integrated Gaussian function is simply determined by the ratio of the standard deviation of the threshold stimulus level to the mean threshold stimulus level.

$$R = \text{Std Deviation} / \text{Mean}$$

If the relative noise level  $R$  is small the node's input-output function will be relatively steep. Verveen [4] found that the steepness of the "best-fit" integrated Gaussian function was a strong function of the fiber's nodal surface area, expressed as a function of nodal diameter. For smaller diameter fibers, relatively large changes in stimulus amplitude were necessary to cause a given change in discharge probability. The value of  $R$  determines the steepness of the integrated Gaussian function. Verveen found the following relationship between fiber diameter  $D$  in microns and the relative noise level  $R$ :

$$R = 0.03 D^{-0.8}$$

This equation was derived from measurements of invertebrate myelinated fibers having node diameters ranging from microns to hundreds of microns using single short-duration pulses.

We expand Verveen's equation to express nodal surface area in terms of nodal length and diameter and recompute the scaling constant for Verveen's original equation assuming 2.0 micron nodal

lengths for the fibers originally studied by Verveen [3].

$$R = 0.052 (L D)^{-0.8}$$

where L and D are in microns.

#### ANATOMY

Lieberman and Oliver [2] measured the diameter and length of the nodes of Ranvier of Type I afferent neurons in the cat cochlea. These neurons are typically bipolar in shape with thin peripheral dendrites and thicker centrally-projecting axons. Anatomical data for these cat neurons are summarized in Table I.

#### PHYSIOLOGY "PREDICTED" FROM ANATOMY

These anatomical data were applied to our modified Verveen model to predict probability-of-firing characteristics for electrically stimulated primary afferents in the cat cochlea. Such predictions can be directly compared with the electrophysiological measurements of Javel et al [1]. Model predictions are presented in Table II. For the purpose of this paper, model predictions include estimates of the R ratios for the integrated Gaussian curves and estimates of the "dynamic ranges" of the nodes. "Dynamic range" is defined as the increase in stimulus amplitude required to cause the probability of discharge to increase from 0.1 to 0.9. Dynamic ranges are expressed in dB.

Recently, Javel et al. [1] recorded VIII nerve single unit responses to electrical stimulation in the cat cochlea. As expected, they found that discharge probability increased when they increased the amplitude of short biphasic pulse stimuli. The rate at which discharge probability increased with stimulus amplitude varied considerably across fibers. In some fibers, as little as a 1 dB increase in stimulus amplitude was required to cause the probability of discharge to increase from 0.1 to 0.9. Other fibers required as much as a 6 dB increase in stimulus amplitude to cause this same increase in discharge probability. "Dynamic ranges" of fibers were evenly distributed over this 1-6 dB range. In other words, there were about the same number of "1 dB fibers" as "3 dB fibers", as "6 dB fibers", etc.

#### DISCUSSION

Predicted dynamic ranges were nearly identical to those directly measured by Javel et al. [1] in the auditory nerve of cat. The agreement between model prediction and measurement is particularly remarkable considering that equations (1) and (2) were obtained from measurements using fibers that are quite different from those in the cat cochlea. Specifically, equations (1) and (2) were derived from measurements of invertebrate myelinated fibers ranging from microns to hundreds of microns.

In this brief report we have illustrated only one example of the predictive power of this stochastic excitation model. This model has proven useful in simulating psychometric functions, intensity discrimination and dynamic range functions.

TABLE I. CAT AUDITORY NEURON ANATOMICAL SUMMARY

	Central Axons (in microns)	Peripheral Dendrites (in microns)
Node Length	1.0	1.0
Node Diameter	0.7 - 2.25	0.1 - 0.7

TABLE II. PREDICTED PROBABILITY-OF-FIRING CHARACTERISTICS FOR CAT NEURONS

	Central Axons	Peripheral Dendrites
R	.027 - .07	.07 - .33
Steepness of I-O curve	steep	shallow
Dynamic Range (dB)	0.6 - 1.56	1.56 - 7.84

#### REFERENCES

1. Javel E, Tong YC, Shepard RK, Clark GM (1987) Responses of cat auditory-nerve fibers to biphasic electrical current pulses. Ann Otol Rhinol Laryngol 96(Supl 128):26-30.
2. Lieberman MC, Oliver ME (1984): Morphometry of intracellularly labeled neurons of the auditory nerve: correlations with functional properties. J Comp Neuro 223:163-176.
3. McNeal DR (1976): Analysis of a model for excitation of myelinated nerve. IEEE Trans Biomed Engin 23:329-337.
4. Verveen AA (1962): Fibre diameter and fluctuation in excitability. Acta Morphologica Neerlandico-Scandinavica 5:79-85.
5. Verveen AA, Derksen HE (1968): Fluctuation phenomena in nerve membrane. Proc of the IEEE 56:906-916.

**A FINITE-ELEMENT MODEL OF BIPOLAR FIELD PATTERNS IN THE ELECTRICALLY  
STIMULATED COCHLEA - A TWO DIMENSIONAL APPROXIMATION**

Charles C. Finley<sup>1</sup>, Blake S. Wilson<sup>1</sup> and Mark W. White<sup>2</sup>

<sup>1</sup> Neuroscience Program, Research Triangle Institute, RTP, NC 27709

<sup>2</sup> Electrical & Computer Engineering Dept, NC State Univ, Raleigh, NC 27695

**INTRODUCTION**

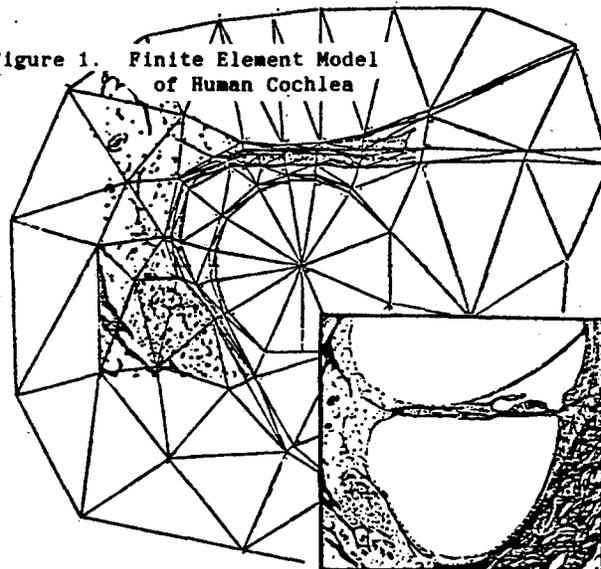
A variety of electric field models have been proposed for the electrically-stimulated cochlea. These models range from closed-form analytical models [2,3] to numerical finite-difference [1,5] models. Of particular interest in these models is a description of the electrical stimulus in the immediate vicinity of the target neural fibers. Present models are limited in their abilities to account for fine details of cochlear anatomy. The analytical approach must assume simplifying geometries and the finite-difference approach is limited in spatial resolution for reasonable problem size. This paper presents initial modeling efforts using the finite-element (FE) approach which allows the description of arbitrary geometries at fine resolution where needed. The model presented here is limited to a two-dimensional cross-section of the cochlea.

**FINITE-ELEMENT MODEL**

In its simplest form the FE approach is a piece-wise linear approximation method in which the field region to be modeled is divided into small discrete elements. Here triangular elements are used with each element being defined by the interconnection of three vertices or nodes. Any one node may define the corner of any number of elements. Potentials within each element are described by planar functions based on the potentials at the element's nodes. Boundary potentials for each side of an element are linear interpolations of the potentials at the nodes which define each side. Regional conductances are defined for each element, and node potentials are either fixed to preset values or left to vary freely. Iterative minimization of the energy in each element of the system (which is equivalent to solving Laplace's equation for each element) produces a field description based on the final potentials at the nodes.

Figure 1 shows a FE description of the cross section of a human cochlea with a bipolar scala tympani electrode pair mounted in a support insulator. The insert shows the histological section represented by the model. The finite element model contains 91 nodes and 161 elements. Node potentials along the outer periphery are fixed at zero. Node potentials at nodes bordering the electrode surfaces are fixed arbitrarily at +100 or -100 mvolts depending on electrode location. Resistivities are defined regionally to

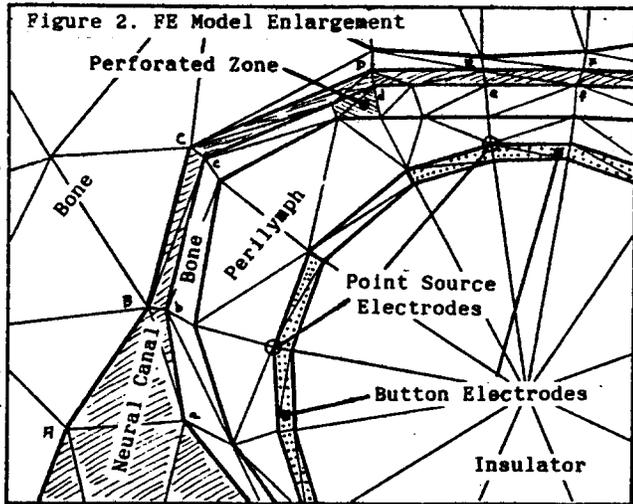
Figure 1. Finite Element Model of Human Cochlea



characterize the electrodes, the carrier insulator, the endolymph, the perilymph, Reissner's membrane, the dendrite tissue leading down from the habenula to the spiral ganglion, the spiral ganglion itself, bone and the pores of the perforated zone beneath the habenula.

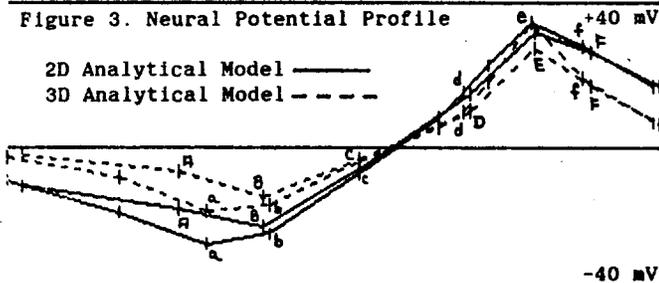
Potential patterns along the neural elements are shown by plotting the potentials of the nodes that lie along the locus of the neural fibers. Figure 2 shows an enlargement of the neural canal and electrode region from Figure 1. Two sets of nodes are used in calculating potential patterns. One is labeled (A-F) and defines the inner modiolar boundary of the neural tissue, whereas the outer scalar boundary is labeled (a-f). Potentials along the loci between nodes are linear interpolations of the node potentials, consistent with the model assumptions.

Limitations of the present model are two fold. First, element selection is not optimal in that some triangular elements approximate narrow isosceles triangles rather than optimal equilateral triangles. This selection reduces accuracy of the field calculations along the long dimensions of the triangles. Second, selection of the two dimensional (2D) geometry makes the implicit assumption that all regions of the cross section extend to infinity both above and below the cross section plane. This may be a useful approximation for the soft tissue structures but is a significant distortion of the electrode surfaces. To assess the impact of this



distortion, two analytical model calculations were performed where the electrodes were reduced to discrete single nodes and modeled first as point dipole sources in homogeneous, three-dimensional (3D) space and then as infinite line charges cutting a homogeneous, 2D plane. Potential plots along the neural tissue loci were generated using the analytical models and the results are shown in Figure 3.

Figure 3. Neural Potential Profile



The results of this comparison are somewhat counter-intuitive. The profiles are generally similar in shape but the gradients for the 3D model are smaller than for the 2D infinitely long electrode. One would have thought that gradients associated with the 3D dipole model would have been steeper. This would clearly have been the case along axes perpendicular to the cross section; however, in the plane of the cross

TABLE I. FE MODEL RESISTIVITY CONDITION SETS

REGION	Regional Resistivity in OHM-CM				
	COND A	COND B	COND C	COND D	COND E
Insulator	70*	$10^4$	$10^4$	$10^4$	$10^4$
Electrodes	70*	1	1	1	1
Scalae	70	70	70	70	70
Reisner's Memb.	70	70	$10^4$	$10^4$	$10^4$
Bone	70	70	420	630	420
Perforated Zone	70	70	420	630	70
Dendrite Canal	70	70	300	300	70
Ganglion Region	70	70	300	300	70

\* Electrodes shrunk to point sources at circled nodes shown in Figure -; otherwise, they are modeled as full button-shaped electrodes.

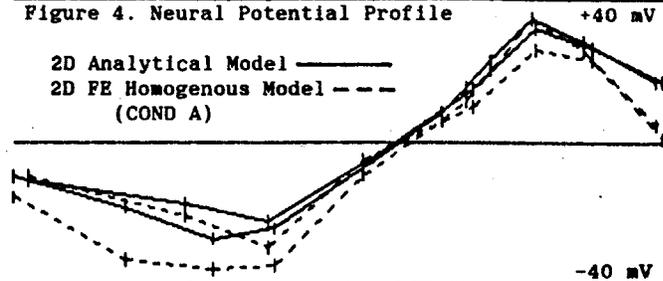
section the 3D gradients along the neural loci are smaller since the 3D field collapses faster than the 2D field. In any case, the 2D approximation used here may tend to over estimate potential gradients along the neural locus.

## RESULTS

To explore the impact of tissue impedance manipulations on the potential fields at the neurons, a series of five model computations are presented. Table I summarizes the regional resistivities selected for each computation.

First is a comparison of the 2D analytical model and the 2D FE model where the electrodes are reduced to point sources (COND A). Here both models assume the same electrode geometry (infinite line) and homogenous tissue region and thus should predict the same results. Figure 4 shows the neural potential fields for both models.

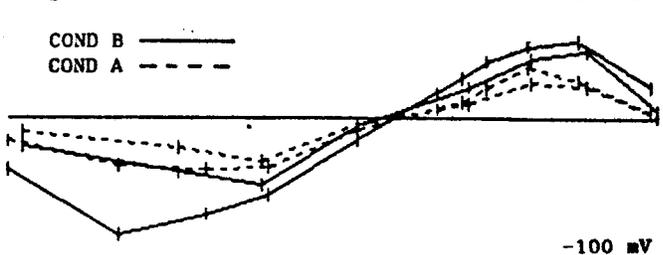
Figure 4. Neural Potential Profile



It is evident that the FE model approximates the 2D analytical solution quite well along the region of the dendrites. However, divergence does occur in the ganglion region and at the tips of the dendrites where element size tends to increase, suggesting that smaller elements should be used in these areas.

In the next comparison, the electrodes are normal size and sit within an insulative carrier. Figure 5 shows the comparison of the line source FE model (COND A) and standard geometry electrodes (COND B) in a homogenous medium.

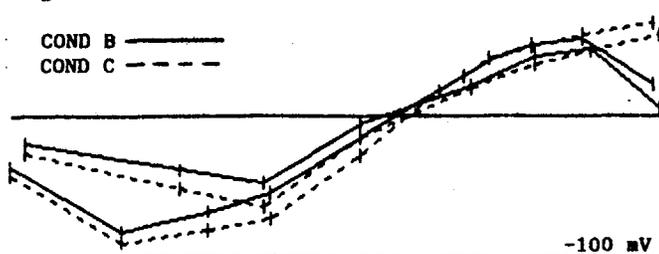
Figure 5. Neural Potential Profile



With the broader electrodes, gradients in regions near the electrodes are small and the conductive edges of the electrodes are closer together. This produces steeper gradients along the dendrites than produced by the line source model.

Next the homogenous tissue medium is modified to reflect realistic tissue impedances. Typical resistivities for bone, neural paths, and perilymph fluids are specified for COND C. Model predictions for homogenous tissue (COND B) and realistic tissue impedances (COND C) are shown in Figure 6.

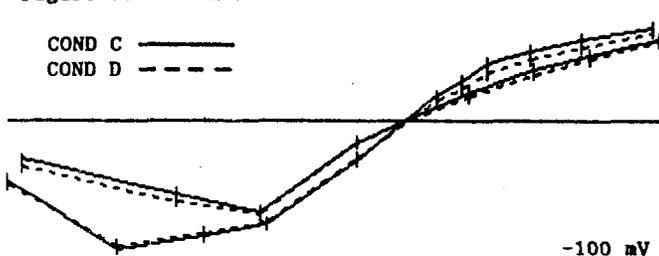
Figure 6. Neural Potential Profile +100 mV



In the region of the dendrites, little difference is seen in the prediction of the models. The abrupt fall off of the fields at the tips of the dendrites for COND C are a consequence of the presence of the high impedance barrier of Reisner's membrane. Also slightly lower potentials values are seen in the region of the ganglion cells.

Recently, Spelman et al. [4] have measured higher bone impedances in the guinea pig on the order of 630 ohm-cm. Figure 7 shows the impact of changing bone impedance from 420 ohm-cm (COND C) to 630 ohm-cm (COND D).

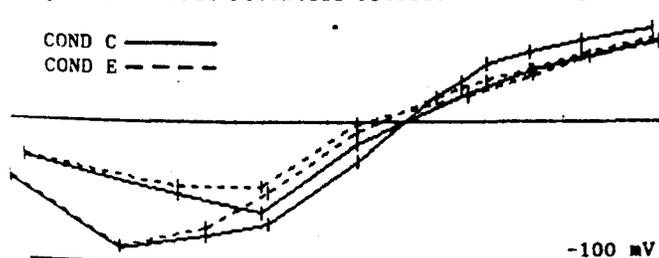
Figure 7. Neural Potential Profile +100 mV



By manipulating bone impedances over this range, little effect is seen on the neural potential profiles. This suggests that for electrodes of this shape and placement, detailed specification of tissue characteristics may not be required to describe neural potential profiles.

The final comparison explores the impact of current shunting through the perforated zone into the neural tissue. This current shunting would be strongest in cases where the neural dendrite tissue was missing, as is the case in many advanced pathologies. COND E assumes that the neural tissue is replaced by highly conductive fluid and that a low impedance path exists through the scala tympani wall beneath the habenula into the dendritic canal. Figure 8 shows the potential profiles for this condition (COND E) compared to the intact cochlea (COND C).

Figure 8. Neural Potential Profile +100 mV



The impact of such a current shunt on the potential profiles is small. Dendrite channel potential profiles are similar, although the gradient across the dendrite channel has been reduced. In the ganglion region, where any surviving elements would remain in the severe pathological condition, field potentials are altered very little. This suggests that current shunting through the perforated zone would alter neural potential profiles very little.

Work is in progress to increase the element resolution of the present model and to expand it to three dimensions. Future model manipulations should include different electrode configurations and placements within scala tympani, as well as calculation of electrode currents.

#### CONCLUSIONS

For a closely-placed bipolar scala tympani electrode:

1. First order effects on neural locus potential gradients are primarily due to electrode geometry.
2. Second order effects on neural locus potential gradients are due to tissue impedance effects.
3. Current shunting through the perforated zone does not alter neural locus potential gradients substantially.

Consequently, design criteria for closely-placed bipolar optimum electrodes should emphasize electrode geometry and placement, with little need to focus on tissue impedance effects.

#### REFERENCES

1. Grizon G (1987): Investigation of current flow in the inner ear during electrical stimulation of intracochlear electrodes. MS Thesis in EE&CS, MIT.
2. Rubinstein JT, Soma M, Spelman FA (1985): Mixed boundary value problem in the implanted cochlea: An analytical model of a cylindrical banded electrode array. IEEE/7th Conf. EMBS, 1120-1123.
3. Soma M, Spelman FA, Rubinstein JT (1984): Fields produced by the cochlear prosthesis: The ear as a multilayered medium. IEEE Frontiers of Eng & Computing in Health Care, pp. 401-405.
4. Spelman FA, Clopton BM (1987): Measurement of the specific impedance of bony tissues in the guinea pig cochlea. Abst. #41 of 10th Midwinter Res Mtg of Assoc for Res in Otolaryngology.
5. Wilson BS, Finley CC (1984): Speech processors for auditory prostheses. 2nd-4th Qtr. Prog Reports, NINCDS, NIH.

This work was supported by NIH contracts N01-NS-3-2356 and N01-NS-5-2396, from the Neural Prosthesis Program.